DISCRIMINATION AND TRANCRIPTIONAL RESPONSE ANALYSIS OF HEMIPTERAN PHYTOPATHOGEN VECTORS

By

SHARON ANDREASON

Bachelor of Science in Biology

University of Texas at Tyler

Tyler, Texas

2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2016

DISCRIMINATION AND TRANSCRIPTIONAL RESPONSE ANALYSIS OF HEMIPTERAN PHYTOPATHOGEN VECTORS

Dissertation Approved:

Astri Wayadande, Ph.D.

Dissertation Adviser

Jacqueline Fletcher, Ph.D.

Francisco Ochoa-Corona, Ph.D.

Ulrich Melcher, Ph.D.

ACKNOWLEDGEMENTS

I would first like to extend my deepest thanks to my advisor, Dr. Astri Wayadande, for all of her years of support. Always full of positivity, encouragement, and advice, her enthusiasm for research and effort devoted to others are attributes that I aspire to cultivate in my future career. I am forever grateful for her guidance throughout earning my degree. I would also like to thank my committee members, Drs. Jacqueline Fletcher, Francisco Ochoa-Corona, and Ulrich Melcher, each of whom have not only offered me thoughtful advice through my research, but have also educated and inspired me in unique and invaluable ways. I am deeply appreciative of their guidance over the years. I want to thank Dr. Judith Brown, without whom I would not have had the opportunity to work on such an important vector system. I am sincerely grateful to Dr. Mohammad Arif, who provided exceptional expertise and advice through my whitefly projects. I would also like to extend my thanks to Dr. Trenna Blagden for all her help and guidance on my projects, particularly during my work with EDNA, as well as for the professional advice she has provided. I want to thank Dr. William Schneider, who welcomed me onto 'Team EDNA' and provided me with excellent opportunities through that work. I also want to thank Jon Daniels for all his help with learning EDNA. Thank you especially to Dr. Brian Couger, whose help with my transcriptome analysis has been absolutely indispensable. Finally, I want to thank my friends and family, especially my parents, both of whom have supported me, encouraged me, and inspired me, and my sister Faith, my bestfriendsister and comic relief. I could not have made it to this day without each of you.

Name: SHARON ANDREASON

Date of Degree: JULY, 2016

Title of Study: DISCRIMINATION AND TRANSCRIPTIONAL RESPONSE ANALYSIS OF HEMIPTERAN PHYTOPATHOGEN VECTORS

Major Field: PLANT PATHOLOGY

Abstract:

Insect vectors play a prominent role in the epidemiology of the pathogens they transmit, often acting as vehicles of pathogen movement into new areas. With the globalization of world economies, pathogen and vector introductions increase; however, the impact can be minimized with use of rapid detection strategies at ports of entry. The development and application of novel, reliable tools that accurately detect pests before they enter the country can facilitate interventions that prevent introductions of exotic invasive species. The work described herein provides new discrimination strategies to this end. *Bemisia tabaci* is an agriculturally damaging whitefly pest species that transmits over 200 viruses to hundreds of susceptible host plants, many of agricultural importance. Compounding the problems caused by *B. tabaci*, is its genetic and biological diversity. As a cryptic species, one with genetically distinct, non-interbreeding variants but morphological homogeneity, highly problematic, non-native biotypes can be distinguished from the native biotype only by molecular tools. As such, three discriminatory techniques were designed for two high-consequence biotypes (B and Q), and the native biotype (A), as well as another whitefly virus vector species, *Trialeurodes vaporariorum*. A reliable multiplex PCR protocol, that incorporates the use of sensitivity-enhancing 5' A/T-rich overhang sequences, was designed and validated by testing over 125 whiteflies from over 55 populations and an inclusivity panel of 17 non-target organisms. Primers were also designed for real-time PCR-based and isothermal amplification discrimination. Melting temperature analysis and helicase dependent amplification protocols were developed and tested for identification of *B. tabaci* B, Q, and A biotypes, and *T. vaporariorum*. Leafhoppers are important vectors of phytopathogenic mollicutes. As pathogens can be carried and introduced undectected via their insect vectors, use of e-probe diagnostic nucleic acid analysis (EDNA) was tested for phytopathogen detection in an insect. The vector *Exitianus exitiosus* harboring *Spiroplasma kunkelii* was used to validate pathogen detection in an insect-derived Next Generation Sequencing dataset background. Finally, the effects of *S. kunkelii* infection in *E. exitiosus* were explored by transciptome sequencing and differential espression analysis. As more is discovered about pathogen and vector interactions, more innovative strategies for pest prevention and management can be devised.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

Table Page

CHAPTER VI

LIST OF FIGURES

Figure Page

CHAPTER III

CHAPTER IV

Figure Page

CHAPTER V

CHAPTER VI

CHAPTER I

INTRODUCTION

Insect vectors play an important role in the epidemiology of many phytopathogenic viruses and all known plant-infecting mollicutes (spiroplasmas and phytoplasmas). Mobile insect vectors allow for the precise acquisition and inoculation of tissue-specific mollicutes and viruses from one plant into another. Insects in the order Hemiptera comprise the majority of phytopathogenic virus and mollicute vectors. Within the Hemiptera, the families Aphididae (aphids), Aleyrodidae (whiteflies), and Cicadellidae (leafhoppers) transmit most of the insect-transmitted plant viruses, and Cicadellid leafhoppers transmit the majority of plant-infecting spiroplasmas and phytoplasmas. These vector families occur worldwide and cause damage that can result in significant yield losses for growers. While the vectors themselves reduce host plant vigor, it is their transmission of pathogens to susceptible hosts which devastates crops, at times causing food limitations in many areas of the world.

As agricultural markets continue to globalize, the importance of having timely and robust agricultural biosecurity programs and capabilities increases along with the risk of pest and pathogen movement and introduction. Because infection with insect-transmitted pathogens could be circumvented by preventing vector introduction and feeding on susceptible hosts, rapid discrimination of vectors and pathogens is essential to maintaining successful agricultural biosecurity efforts. Additionally, enhanced understanding of the molecular interactions between a

phytopathogen and its insect vector can provide new insights into preventing vector-mediated transmission. The work presented here provides new avenues of vector discrimination and pathogen detection as well as increases understanding of microbe-vector interactions in two Hemipteran systems.

The sweet potato whitefly, *Bemisia tabaci* (Hemiptera; Aleyrodidae) (Gennadius), is a globally distributed agricultural pest. It damages plants directly by feeding, reducing vigor by extracting host plant phloem, and by secreting honeydew, a medium for sooty mold growth, that prevents gas exchange and photosynthesis. *B. tabaci* also damages plants indirectly by transmitting plant viruses. The begomoviruses (Geminiviridae) are transmitted by *B. tabaci* alone and account for millions of dollars in crop losses worldwide (Varma and Malathi, 2003).

B. tabaci is a cryptic species, that is, a species composed of several morphologically indistinguishable but genetically distinct, non-interbreeding biotypes (Bellows et al., 1994; Xu et al., 2010). These biotypes differ with respect to host colonization, virus transmission efficiency, pesticide resistance, and invasiveness. Two biotypes are currently resident in the United States, both of which were introduced by trade in the last three decades (Costa et al., 1993; Dennehy et al., 2005). The first introduction led to the displacement of the endemic population with a higher consequence biotype. The second introduced a biotype with close to total pesticide resistance (Horowitz et al., 2005).

Globalization efforts and increased trade among nations escalate the inadvertent introduction of exotic pests and pathogens into the United States (Hulme, 2009). Rapid and accurate diagnostic tools are essential for successful interception and elimination of these pests. Molecular diagnostic techniques are used extensively in the diagnosis of plant diseases. These methods can be applied to the identification of morphologically indistinguishable or variable agricultural pest species, such as the *B. tabaci* sibling species group. Tools based on molecular and genomic characterization of agricultural pests can help prevent the inadvertent introduction of exotics.

While the introduction of a foreign insect vector is problematic, the establishment of such a vector harboring an exotic pathogen is of greater concern. Records at United States ports of entry show that from 1984 – 2000 nearly 35,000 insect pests on average were intercepted each year, and there is an increase in this trend (McCullough et al., 2006). Of these interceptions, insects in the order Hemiptera accounted for the highest percentage of pests seized at the borders, at nearly 37%. These numbers suggest that there exists great potential for the accidental introduction of hemipteran vectors. There is a need not only for rapid discrimination of vectors but of the phytopathogens they could be harboring.

The ability of an insect to acquire and subsequently transmit a pathogen into a plant is a complex and evolved one. While the capability to quickly detect and discriminate these vectors and their pathogens is essential to making various quarantine or management related decisions, an understanding of this complex process at the genetic level could significantly enhance control options as well. With the advent of next generation sequencing, genome and transcriptome research has been revolutionized (Mardis, 2008). The availability of this technology for insect vector research provides vast opportunities to elucidate the transmission process more clearly, particularly at the molecular and genetic levels. By comparing the transcriptomes of plant pathogen-harboring and pathogen-naïve vectors, transcripts that are differentially expressed in response to pathogen acquisition can be discovered. This information can in turn be used for identifying and testing targets for manipulating or preventing transmission, such as in a leafhopper-spiroplasma vector system.

The gray lawn leafhopper, *Exitianus exitiosus* (Uhler) (Hemiptera: Cicadellidae), is a ubiquitous Poaceae (grass) pest throughout the United States and vector of two pathogens, *Maize chlorotic dwarf virus* and *Spiroplasma kunkelii* (Nault and Madden, 1988; Nault, 1980). *E. exitiosus* was first recorded as a crop pest in 1879, when it was reported to have damaged wheat production in some areas of the Eastern United States. It transmits *S. kunkelii* in a propagative manner; that is, after the mollicute is ingested, it traverses the gut epithelium to enter the hemolymph, where it then moves

into various organs and multiplies, and finally it enters the salivary glands where it is egested into the plant with vector saliva (Purcell, 1982).

The corn stunt spiroplasma, *S. kunkelii*, is an obligate parasite of *Zea* species, causing yellowing, stunting, reddening of leaf margins, and yield losses in its teosinte hosts. *S. kunkelii* is an important pathogen of maize in the Americas (Bradfute et al., 1981). Some resistant cultivars have been developed; however, they have become unstable over time, suggesting that the corn stunt spiroplasma can successfully overcome host resistance (Silva et al., 2003; Carpane et al., 2013). Many management strategies are then focused on vector control. Therefore, corn stunt disease management options can be enhanced greatly as more is understood about the vector transmission of its causal pathogen.

The objectives of this research are to provide new protocols for rapid discrimination of insect vectors, detect plant pathogens within vector-derived next generation sequencing datasets, and identify differentially expressed transcripts within insect vectors in response to phytopathogen acquisition:

1. Design discriminatory multiplex PCR, helicase dependent amplification, and melting temperature analysis protocols for rapid determination of high-consequence *B. tabaci* biotypes.
2. Validate the use of EDNA (E-probe Diagnostic Nucleic acid Analysis) for rapid detection of *S. kunkelii* within *E. exitiosus* transcriptome sequence datasets.
3. Evaluate and compare the transcriptomes of *S. kunkelii*-infected and *S. kunkelii*-naïve *E. exitiosus* to identify transcripts that are differentially expressed in response to pathogen acquisition.

CHAPTER II

REVIEW OF LITERATURE

Agricultural Biosecurity

Agricultural biosecurity encompasses the measures taken to protect a country, region, farm, etc. from the introduction of foreign pests and pathogens that have the potential to be damaging to agricultural products (Waage and Mumford, 2008). It includes the policies and regulations enacted to prevent such an event as well as the expertise and technologies employed to monitor and diagnose incoming pests. Both the illegal, targeted introduction of pathogens and the accidental, natural introduction of pests and pathogens are addressed in agricultural biosecurity.

Since the terrorist attacks of September 11, 2001 and the following anthrax bioterrorist acts, more attention and emphasis has been placed on the intentional, harmful introduction of human, animal, and plant pathogens than previously as policy makers and the public were made aware of the United State's vulnerability to such organisms (Meyerson and Reaser, 2002). While an intentional introduction can pose a threat to the functionality and productivity of agriculture, an unintentional introduction can have equally damaging consequences. With the current globalization effort and consequential increase in international trade, there has become greater potential for the inadvertent introduction of exotic pests and plant pathogens into countries where

they have not previously occurred (Hulme, 2009). Increased import and export of agricultural products and ornamental plants results in greater potential for the transfer of plant pests and pathogens between countries.

The introduction of the insect vector of *Candidatus* Liberibacter spp., the Asian citrus psyllid *Diaphornia citri* (Kuwayama), to Southern Florida in 1998 raised concerns for the establishment of huanglongbing (HLB) citrus greening disease in Florida (Halbert and Manjunath, 2004). At this time, the establishment of both insect vector and pathogen of one of the most consequential insect-transmitted diseases of citrus, which could result in millions of dollars in losses for the Florida citrus industry, was highly concerning. Response to the vector introduction resulted in the design of Liberibacter diagnostic tools, which have been used to monitor for pathogen presence. Subsequently, HLB was discovered in Florida in 2005 and threatens the future of citrus production in the U.S. (Gottwald et al., 2007). Preventing the introduction of the vector could have eliminated the disease threat altogether. Therefore, preventing the introduction of the pathogen vector is fundamental to preventing disease spread and/or initial introduction of the pathogen. This logic can be applied to *B. tabaci* biotypes and the plant viruses it transmits.

Bemisia tabaci

History and Distribution

The sweetpotato whitefly, *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae), was first named and described when it was discovered in tobacco plantations in Greece in 1889 (Gennadius, 1889). It was then recognized as an agricultural pest because it damaged tobacco plants upon which it was feeding. In the United States, *B. tabaci* was first collected in 1897 on

sweet potato and was subsequently given its common name, the sweetpotato whitefly, by Quaintance (1900). Over the following 60 years, numerous whitefly species were described and named. However, in 1957, confusion over the nomenclature of seemingly identical whitefly species found on different hosts led to the reclassification of the species (Russell, 1957). Eighteen of the then described whitefly species were identified as the same species and classified into the taxon *B. tabaci.* Its members were then termed biotypes, determined by their host ranges and differential transmission of plant viruses.

The country of origin of *B. tabaci* is unknown, though it is thought to be India or Pakistan as these countries harbor the greatest diversity of *Bemisia* parasitoids, a characteristic which is often indicative of the point of origin of a species (Mound and Halsey, 1978). *B. tabaci* was previously documented in primarily tropical and subtropical regions (Cock, 1986). These warm winter climates allow for survival of the species in outdoor crops. However, *B. tabaci* is now established on all continents except Antarctica (De Barro et al., 2010). This nearly global distribution is the result of the trade of ornamental crops between 1985 – 1986 when the *B. tabaci* B-biotype was unintentionally transported to North and Central America (U.S. and Mexico), South America (Venezuela, Columbia, Brazil, and Argentina), the Caribbean, Europe, the Mediterranean Basin, Africa, and Asia (Brown et al., 1995). Protected greenhouse environments are responsible for their persistence in cooler temperate regions (De Barro, 1995). Both indigenous and invasive populations are now established worldwide, and *B. tabaci* is considered one of the most damaging pests globally (De Barro et al., 2011). The origin, dispersal, establishment, and interactions of *B. tabaci* biotypes are discussed in detail in a later section.

Biology

B. tabaci is classified in the order Hemiptera, suborder Sternorrhyncha, and family Aleyrodidae. These insects are approximately 1 mm in length with a pale yellow body and white

wings (the latter due to the deposition of a powdery wax) (Basu, 1995). Whiteflies are haplodiploid (Byrne and Devonshire, 1996); females are diploid as a result of fertilization of the egg, and males are haploid, resulting from a lack of fertilization. They are also multivoltine with 11 – 15 generations produced per year in favorable tropical, subtropical, or fringe-temperate conditions (Butler et al., 1983). Females lay between 50 – 300 eggs in their lifetime depending on environmental conditions, and the period of development from egg to adult can range from 11 – 14 days in optimal conditions or warmer months to 6 – 12 weeks in winter months (Basu, 1995). Newly hatched crawlers, or the first instar, crawl on the leaf surface to seek out a vascular bundle for feeding (Cohen et al., 1996, 1998). After locating a suitable feeding site, the crawler molts and becomes sessile until it reaches maturity.

Feeding by both immature and adult whiteflies is characteristically on the undersides of leaves of mostly herbaceous, annual plants (Mound, 1983). *B. tabaci* is primarily polyphagous in nature, feeding on over 500 plant species in 74 different families (Mound and Halsey, 1978), including economically important vegetable crops, grains, legumes, cotton, and ornamentals (De Barro et al., 2010). However, some biotypes, such as the Jatropha race in Puerto Rico, have been shown to be monophagous. Whiteflies feed on phloem using piercing-sucking mouthparts by inserting their stylets intercellularly through plant epidermal and parenchymal tissue until they reach the phloem wherein stylet penetration becomes intracellular (Pollard, 1955; Costa, 1969; Janssen et al., 1989). The time required to reach the phloem is 30 minutes on average but at times, especially after starvation, can be 15 minutes (Capoor, 1948; Walker and Perring, 1994). The saliva secreted during feeding polymerizes to form a stylet sheath in the host plant tissue, resulting in a track of the direction and path of stylet penetration (Cohen et al., 1998). The uptake and digestion of sugary phloem sap results in the excretion of honeydew by the whitefly, which serves as a means of maintaining the osmotic gradient between the gut lumen and hemolymph (Fisher et al., 1984).

The advent of electronic penetration graph (EPG) analysis of probing behaviors by piercing-sucking hemipterans allowed for more in depth investigation and understanding of how whiteflies feed on plants (McLean and Kinsey, 1964; Tjallingii, 1978; Janssen et al., 1989). By connecting a wire from an AC (or DC) circuit monitor output source to a leaf, and an input wire to the insect, an electrical circuit is completed upon the insect's contact with the leaf. The monitor reads changes in the current that result from different activities (stylet insertion, ingestion of phloem, oviposition, etc.). EPG was first used to study feeding behavior of the greenhouse whitefly, *Trialeurodes vaporariorum*, by Janssen et al. (1989), then of *B. tabaci* by Walker and Perring (1994). By comparison with EPG results for aphids and *T. vaporariorum*, feeding and oviposition by *B. tabaci* biotype B (*B. argentifolii* Bellows and Perring) are very similar (Walker and Perring, 1994). Stylet penetration was represented by large-amplitude sawtooth waveforms occurring as the leaf epidermis, mesophyll, and vasculature were traversed intercellularly. Penetration was followed by a sudden, short drop in voltage, resulting in a transition waveform, identified as intracellular penetration of the phloem (Jiang et al., 1999), and then ingestion of phloem sap resulted in high-flat waves. EPG experiments on whiteflies were later expanded to study the differences in feeding behavior on hosts versus non-hosts (Walker and Perring, 1994; Walker, 1997; Lei et al., 1998), between the *B. tabaci* B and Q biotypes (Jiang et al., 1999), and in transmission of *Tomato yellow leaf curl virus* (TYLCV) by *B. tabaci* (Jiang et al., 2000), all of which will be discussed in more detail in a later section.

Plant Damage

B. tabaci damages plants both directly and indirectly. Feeding by the whitefly reduces host plant vigor as it extracts the phloem sap, which contains nutrients essential to the plant's health (Berlinger, 1986). This can become more problematic in cases of high whitefly infestation than when a relatively low number of insects are feeding on the plant. The secretion of honeydew

also contributes to the whitefly-induced demise of a host plant. This sticky substance can cover stomata, blocking gas exchange and preventing photosynthesis (Fisher et al., 1984; Berlinger, 1986). Again, damage of this kind worsens as the level of whitefly infestation increases. Direct damage by *B. tabaci* is relatively inconsequential, however, when compared to its capacity for plant virus transmission.

Virus Transmission

B. tabaci transmits plant viruses in five families, the Betaflexiviridae, Closteroviridae, Potyviridae, Secoviridae, and Geminiviridae (Jones, 2003). In the family Betaflexiviridae, *B. tabaci* transmits two carlaviruses, *Cassava brown streak virus* and *Cowpea mild mottle virus*. Both viruses are transmitted non-persistently by the whitefly vector (Jeyanandarajah and Brunt, 1993) and constitute the only viruses known to be transmitted in this manner by *B. tabaci*. Criniviruses in the family Closteroviridae are transmitted by whiteflies and mealybugs in a semi-persistent, non-circulative manner (Jelkman et al., 1997; Agranovsky, 1996). Several viruses, such as *Sweet potato mild mottle virus* and *Squash yellow leaf curl virus*, in the family Potyviridae, also are whitefly-transmitted. These flexuous, filamentous rods are transmitted to host plants by *B. tabaci* in a semi-persistent manner as well (Hollings et al., 1976; Muniyappa and Reddy, 1983; Zouba and Lopez, 1998). Members of the family Secoviridae, including *Tomato torrado virus*, also are transmitted by whiteflies, though the mode of transmission is not yet well understood (Amari et al., 2008).

The vast majority of the viruses known to be transmitted by *B. tabaci* are in the genus *Begomovirus*, family Geminiviridae. These are generally of the most concern to researchers working on *B. tabaci*-transmitted viruses as begomoviruses represent over 90% of known whitefly-transmitted viruses (Jones, 2003; Wintermantel, 2004). *B. tabaci* is the only known vector of begomoviruses, transmitting over 200 viruses in this taxon alone (Frohlich et al., 1999;

Brown, 2001; Brown and Czosnek, 2002). These geminiviruses have a moderate to broad host range, infecting dicotyledonous plants including many economically important crops such as cotton, cassava, tomato, potato, and bean. Crop losses resulting from begomovirus infection can range from 20% to 100% depending on the crop and cultivar, growing season, and plant stage at infection.

Begomoviruses are composed of circular, single-stranded DNA, a trait which is relatively uncommon among plant viruses (Goodman, 1977; Reisman et al., 1979). They, like their geminivirus family, have either a monopartite or bipartite genome organization encompassed in a geminate icosahedral capsule. Begomoviruses that are native to the Old World are both monopartite and bipartite, whereas those that originate in the New World are only bipartite. The DNA-A component is 2.6 kb in length and includes five open reading frames, two of which are highly conserved and encode replication and coat proteins (Hanley-Bowdoin et al., 1999). The conserved nature of the coat protein (>70% amino acid similarity worldwide) is significant as it is a determinant for transmission by *B. tabaci* (Harrison and Murant, 1984). The DNA-B component of the bipartite genome (also 2.6 kb in length) encodes two proteins that facilitate systemic movement and influence host range (Sanderfoot and Lazarowitz, 1996).

Mode of Transmission

There are four described interactions between plant viruses and an insect vector's body; these include non-persistent transmission, semi-persistent transmission, circulative transmission, and propagative transmission (Watson and Roberts, 1939; Kennedy et al., 1962; Sylvester, 1958; Nault, 1997). Research on the transmission pathway of luteoviruses in the aphid vector has contributed a model for circulative transmission in insect vectors (Gray and Gildow, 2003). Evidence gathered from transmission studies in *B. tabaci* supports the conclusion that, like luteoviruses, begomovirus transmission follows a circulative pathway (Cohen and Nitzany, 1966;

Rubinstein and Czosnek, 1997). In this interaction, virions are acquired from infected plant material by the feeding insect and move into the foregut. Rather than being transmitted from the mouthparts or foregut as the insect salivates, as with non- and semi-persistently transmitted viruses respectively, the virions then move into the midgut and hindgut (Gray and Banerjee, 1999). Here they must cross from the gut lumen into the epithelial cells by endocytosis, traverse the gut tissues, and exit gut cells by exocytosis to enter the hemocoel (Garret et al., 1993). Within the hemocoel, virus particles migrate via hemolymph to the salivary glands and salivary duct, where they can be inoculated into a plant during subsequent feeding.

The data collected to date on the transmission of begomoviruses in the whitefly vector lack the comprehensiveness of that of luteoviruses in aphids; however, some interesting facets have been discovered. An EPG study of *B. tabaci* in the transmission of TYLCV revealed that a short $E(pd)_1$ waveform was present in 94% of inoculation tests that successfully established disease in test tomatoes (Jiang et al., 2000). In comparing similar EPG results for virus transmission in aphids, the $E(pd)_1$ waveform was suggested to be salivation into the phloem sieve elements by Lei et al. (1998) and Jiang et al. (1999), and the results of the Jiang et al. 2000 study were consistent with this observation. Phloem salivation is then a mechanism by which phloem-restricted plant viruses are inoculated into host plants by insect vectors.

Several studies have reported similar acquisition access period (AAP) and inoculation access period (IAP) requirements for transmission of different begomoviruses (Brown and Czosnek, 2002). Minimal AAPs for several isolates range from 15 to 60 minutes, while minimal IAPs are 15 – 30 minutes. Different AAP and IAP combinations result in varying transmission efficiencies, and efficiency can vary among begomovirus and *B. tabaci* biotype vector combinations. An average latent period (LP) of 6 – 12 hours is required for the virions to traverse the vector body, moving across membrane barriers (Bird and Maramorosch, 1978). In another

study, the general pathway and temporal translocation of TYLCV in *B. tabaci* was traced by performing PCR and immunocapture-PCR on dissected adult head, midgut, hemocoel, and salivary glands (Ghanim et al., 2001). Both methods verified an approximate 8 hour latent period from acquisition to inoculation.

Studies using PCR for detection have shown that the greenhouse whitefly, *Trialeurodes vaporariorum*, a vector of criniviruses, is capable of ingesting begomovirus virions (*Squash leaf curl virus* (SLCV) and TYLCV). However, this species is unable to transmit them (Polston et al., 1990; Antignus et al., 1993; Rosell et al., 1999; Czosnek et al., 2002). Rosell et al. 1999 failed to detect by PCR the SLCV coat protein gene in challenged *T. vaporariorum* saliva and hemolymph samples. Further, Czosnek et al. showed that TYLCV was undetectable by PCR in the hemolymph and salivary glands, even after an AAP of 24 hours. The results of both studies suggest an inability of begomovirus virions to cross from the gut to the hemolymph in *T. vaporariorum*. This virus-vector specificity is determined by the begomovirus coat protein (CP) and (presumably) receptors in the *B. tabaci* gut epithelium (Briddon et al., 1990; Azzam et al., 1994; Hofer et al., 1997). *Abutilon mosaic virus* (AbMV) is a begomovirus species that has lost its ability to be transmitted by *B. tabaci* (Wu et al., 1996). The virions collect in the digestive tract, but are absent from the hemolymph (Morin et al., 2000). However, when the CP of AbMV was replaced with that of S*ida golden mosaic virus*, a transmissible begomovirus, the chimeric AbMV was transmitted to host plants (Hofer et al., 1997). Analysis of AbMV coat protein and subsequent mutagenesis of the protein by replacing three key amino acids reestablished full transmissibility (Honle et al., 2001).

The key difference between circulative and propagative viruses is that circulative viruses do not replicate in the insect vector. This has held true for all begomoviruses studied except in a study by Ghanim et al. (1998) using *Tomato yellow leaf curl virus* from Israel (TYLCV-IS). The

authors provided evidence for the transovarial transmission of TYLCV-IS in *B. tabaci* by southern blot hybridization, transmission tests, and by amplifying viral DNA by PCR in eggs, crawlers, and first and second generation adult progeny of viruliferous flies. Though these results have not been successfully replicated, evidence for the passage of begomovirus DNA to whitefly progeny was published by Bosco et al. (2004) as well. They found that *Tomato yellow leaf curl Sardinia virus* (TYLCSV) DNA and not TYLCV (different isolate used) was detected in the eggs, nymphs, and some adults of first generation progeny of infective B biotype (and not Q); however, these progeny did not transmit the virus to test host plants. Therefore, TYLCSV DNA was transmitted transovarially, but infectivity to plants was not demonstrated.

In aphids, the proteins of endosymbiotic microorganisms have been identified as interacting with plant viruses, such as luteoviruses, and are required for their transmission (van den Heuvel et al., 1994). One such protein discovered is the *Buchnera* sp. encoded GroEL homologue (van den Heuvel et al., 1994; Filichkin et al., 1997). These endosymbionts, located in mycetocytes, are found in several insect groups, including whiteflies (Buchner, 1965). *B. tabaci* contains two endosymbionts in mycetocytes: a predominant microorganism that is unrelated to *Buchnera* (Costa et al., 1995; Baumann et al., 1993) and another that is homologous to the aphid endosymbiont (Baumann et al., 1993; Clark et al., 1992). Morin et al. (1999) found that this *Buchnera*-like endosymbiont also produces a GroEL homologue that is required for successful traversal of TYLCV through the whitefly vector. They further showed that GroEL binds to the coat protein of this begomovirus, as well as non-transmissible AbMV, thereby protecting it from degradation by vector defenses in the hemolymph (Morin et al., 2000). The non-transmissibility of AbMV is presumably due to its inability to traverse the epithelial layer of the gut into the hemolymph.

Much remains to be discovered and understood about the transmission of plant viruses by *B. tabaci*, especially of begomovirus-*B. tabaci* relationships. While the general path of transmission is comparable to that of luteoviruses in aphids, the cellular and molecular mechanisms conferring transfer are largely unknown. Difficulty in elucidating whitefly-mediated begomovirus transmission has largely been due to complications in attempts to visualize the virions in the vector. However, some progress was made when Hunter et al. (1998) used indirect-fluorescent-microscopy to detect virions, showing that the filter chamber and anterior midgut were probable sites of passage from the gut lumen to the hemolymph. Recent anatomical studies of *B. tabaci* B biotype by Cicero and Brown (2011a; 2011b) provided insights regarding organ function and involvement in transmission. An interesting observation made by Cicero and Brown (2011a), confirmed by dissections, is that the midgut can translocate through the petiole between the thorax and abdomen. An implication drawn from this is that, given the right positioning, virions may potentially pass directly from the midgut to the salivary glands. Virions were detected by in situ hybridization in the primary salivary glands (PSG); however, consistent with other studies, staining and dissections of the accessory salivary glands (ASG) localized no virions (Cicero and Brown 2011b). Despite this lack of evidence, it is still assumed that the ASGs are involved with transmission, as with aphids, and attempts to confirm this are ongoing. Though begomovirus virions have been detected in the whitefly vector, they have yet to be visualized.

Biotypes

In 1957, two separate populations of morphologically similar *Bemisia tabaci* were observed in Puerto Rico (Bird, 1957). One population was found to be monophagous, feeding only on *Jatropha gossypifolia*, and the other population was polyphagous. Following this observation, the concept of biotypes or host-related races composing *B. tabaci* was suggested. Bird and Maramorosch (1978) observed that these populations also differed in the number of

begomoviruses they transmitted. The Jatropha race transmitted only one begomovirus and the polyphagous Sida race transmitted several begomoviruses. The idea of morphologically identical but biologically differentiated sibling species, or biotypes, became widely accepted in the late 1980s when some *B. tabaci* in the southern United States behaved differently from the indigenous populations. The new population had a different host range than the native one and was therefore called the B biotype, while the native population was called the A biotype (Costa and Brown, 1991). Esterase profiles of the variants further differentiated the two populations genetically. Additional molecular analyses of populations using methods such as allozyme electrophoresis, randomly amplified polymorphic DNA-PCR (RAPD-PCR), restriction fragment length polymorphisms (RFLP), and amplified fragment length polymorphisms (AFLP) further established the evidence of genetic variation among populations of *B. tabaci*. Further, Bellows et al. (1994) described *B. tabaci* biotype B as a distinct species, *B. argentifolii* (Bellows and Perring), based on crossbreeding experiments, mating behavior differences, different allozyme frequencies, genomic DNA variation, and morphological differentiation. Despite the evidence for or against species differentiation, the dispute over whether *B. tabaci* is a complex species or species complex has remained unresolved.

Since the invasion of the B biotype in the U.S. and its global dispersal by the trade of ornamentals in the 1980s, much focus has been placed on understanding the differences among biotypes at biological, ecological and genetic levels. As a species, *B. tabaci* is highly polyphagous; studies focused on the host range of the B biotype attest to this observation (Brown et al., 1995). However, monophagous and oligophagous biotypes have been reported, such as the N biotype (Jatropha race) in Puerto Rico and the Cassava biotype found in Ivory Coast among others (Brown et al., 1995). Considering differences in the host ranges of biotypes, certain biotypes have an increased potential to become invasive in other countries, given the opportunity. The B biotype has accomplished this, having become established in the United States (Costa et

al., 1993). In the last decade, the Mediterranean Q biotype (pesticide resistant) has become invasive to many countries and was recently found the United States (Dennehy et al., 2005).

Resistance to pesticides is another characteristic that varies among biotypes. A comparison of resistance in the B and Q biotypes showed that the latter is resistant to pesticides used to control the former (Horowitz et al., 2005). The application of these pesticides would lend the Q biotype a competitive advantage over the B biotype. However, considering other biological factors, such as fecundity and competitiveness, the B biotype had a greater number of progeny and was able to outcompete and displace the Q biotype in mixed colonies (Pascual and Callejas, 2004).

Geminivirus transmission by *B. tabaci* varies somewhat among biotypes (Bedford et al., 1994). A few years after the introduction of the B biotype to the southern U.S., solanaceous and cruciferous crops that were not previously considered hosts became infected with whitefly-transmitted begomoviruses (Brown et al., 1995). The cause was concluded to be due to a difference in vector biology rather than the introduction of an Old World begomovirus harbored by the introduced B biotype (Brown, 1994). The invading biotype had a larger host range and was able to acquire and transmit New World begomoviruses. Subsequent studies have found that not all biotypes are able to transmit all begomoviruses. The fundamental requirement for potential transmission is the biotype's ability to feed on a particular plant species rather than incompatibility between the virus species and host plant or virus and vector biotype (Bedford et al., 1994). Variability in transmission efficiencies among different virus-vector combinations also exists, but the differences are less consequential than vector-host compatibility.

Biotype Discrimination

Typically the fourth larval instar (also called the pupal stage, through whiteflies do not undergo complete metamorphosis) is examined for the identification of whitefly species, and the defining characteristics are reliable between species (Hodges and Evans, 2005). The morphological identification of *B. tabaci* biotypes, however, presents a greater challenge. There is little to no distinct morphological difference among the biotypes such that no morphological characteristic can be used reliably for identification. Further, the structure of the puparial stage displays high variability dependent on the host plant species and leaf surface on which it develops (Bedford et al., 1994). Thus, the differentiation of *B. tabaci* biotypes must be dependent on molecular methods.

The use of esterase profiles, RFLP, and AFLP, among other methods, have established the genetic variation among biotypes, and these tests have been used for biotype identification (Costa and Brown, 1991; Lee et al., 2000; Cervera et al., 2000), but the use of DNA barcoding techniques for *B. tabaci* has permitted the most rapid identification of biotypes. The mitochondrial cytochrome oxidase I (*mt*COI) gene, identified as a suitable locus, and has become the standard region for barcoding of animals, insects, and many other eukaryotes because it is conserved within a species but exhibits interspecific variation. Within the *B. tabaci* sibling species complex, however, genetic variation in the *mt*COI gene among biotypes is high enough to be used for the identification of *B. tabaci* to biotype. Use of the maternally inherited mitochondrial DNA for identification assumes that crossbreeding, or hybridization, of biotypes does not occur and that the biotypes are reproductively isolated. Crossbreeding studies have shown that hybridization of biotypes does not occur and that there are pre- and post-zygotic barriers that culminate in the death or reproductive failure of hybrid F1 progeny (Ma et al., 2004; Khasdan et al., 2005; Elbaz et al., 2010). Mitochondrial COI sequences can therefore be used to

design biotype-specific PCR primers for rapid discrimination of *B. tabaci* biotypes and provided the foundation of the diagnostic work with this species and *T. vaporariorum* described herein.

Next Generation Sequencing

With the advent of next generation sequencing (NGS), highly improved avenues for genomic and transcriptomic research became possible. Before the development of NGS, sequence-based genome mapping was limited to sequencing of genome fragments using Maxam-Gilbert chemical method then Sanger chain-termination biochemistry. Sanger sequencing, based on the chain-termination method, remained the primary mode of DNA or cDNA sequencing since its invention in 1977 until 2005 when NGS platforms were made commercially available (Sanger et al., 1977; Shendure and Hanlee, 2008; Mardis, 2013). Shotgun sequencing, advanced during the Human Genome Project (Venter et al., 2001), provided the foundation of NGS by allowing for determination of DNA sequences much longer than those (up to approximately 1,000 nucleotides per reaction) achieved by Sanger sequencing. NGS platforms are based on the methodology of shotgun sequencing in that long DNA strands are randomly fragmented, sequenced by synthesis in parallel, and assembled *de novo* (when a reference genome is unavailable) using bioinformatics to determine the full sequence. This technology allows for high-throughput DNA sequencing using condensed, parallel platforms to sequence up to 100 million short reads in a single procedure. Several commercially available NGS platforms are currently in use, including Roche 454 pyrosequencing, Ion Torrent, Illumina Solexa, and Pac Biosciences, among others.

Metagenomics

The capability of NGS to sequence millions of reads per run has allowed for the advancement of the field of metagenomics. Metagenomics encompasses the study of genomes within a community of organisms. Before the advent of NGS technologies, surveys of microbiota in environmental samples were limited to culturable organisms (Schloss and Handelsman, 2005). The inability to isolate and cultivate certain organisms limited the detection and identification of all organisms in a complex sample. With the vast majority of eukaryotic organisms hosting a community of endosymbiotic microorganisms, many of which are fastidious, the ability to sequence unculturable microorganisms provides a significant improvement to metagenomic studies.

This newfound ability to sequence all organisms within a given sample provides a significant opportunity to the realm of diagnostics (Mardis, 2008). Specifically, for plant pathology, including agricultural biosecurity, this allows for detection of any and all pathogens that could be infecting a plant (Adams et al., 2009). By extracting total DNA or RNA from a plant sample and sequencing using an NGS platform, theoretically all organisms within the plant, including any pathogens, can be sequenced and subsequently identified. NGS can then be employed as a single tool capable of detecting different pathogen classes, including bacterial, fungal, viral, or otherwise, in a given sample.

Many effective detection tools and techniques are currently available and continually in development. Immunologically-based diagnostic assays, such as enzyme linked immunosorbent assays (ELISA), are extremely specific, sensitive, and rapid. These tools are very effective at quickly identifying specific viruses. PCR methodologies (real-time (qPCR), multiplex, immunocapture (IC-PCR), etc.) include extremely effective tools for detecting and discriminating specific pathogen species, strains, haplotypes, races, and so on. These techniques can be

developed and adapted to be highly specific and sensitive. Other nucleic acid based discrimination methods, such as AFLP, RFLP, and sequence characterized amplified region (SCAR), additionally have been developed and employed in the detection of plant pathogens and diagnosis of disease. These methods each can precisely and accurately discriminate specific pests and pathogens; however, they all require much prior characterization of the target organism, whether immune-related (e.g. antibody development) or genomic (e.g. primer design), and adaptational, such as fine-tuning primer annealing temperature, specific reagent concentrations, and primer stability. Therefore, these methodologies are limited in that they are developed and used on a case by case basis; that is, no single method can be streamlined to detect and classify any and all target organisms in one procedure. NGS methods can be adapted to accomplish just that.

E-probe Diagnostic Nucleic Acid Analysis

Metagenomic and metatranscriptomic datasets generated by NGS can be, by today's standards, extremely large, on the scale of terabytes of raw sequence data (Richter and Sexton, 2009). Further, this data is initially unassembled, and assembly of the data into annotatable sequence reads takes time. Currently, the time requirement for sequence assembly and annotation is not conducive to the time-sensitive, rapid diagnostic needs of U.S. ports of entry and other agricultural biosecurity entities. Therefore, to address this limitation and adapt NGS to diagnostic resolutions, the e-probe diagnostic nucleic acid analysis (EDNA), a bioinformatic tool that detects target sequences in an NGS dataset while eliminating the need for assembly, was developed by researchers at OSU and the USDA ARS (Stobbe et al., 2013).

EDNA is a novel assay that essentially ignores non-target sequence reads in a metagenomic and metatranscriptomic dataset, thereby allowing for the *in silico* identification of target sequences and subsequent detection of a target organism in a given sample (Stobbe et al., 2013). In developing and conducting an EDNA, electronic probes or e-probes were designated for a target organism and used to query the dataset for presence at a certain E-value. *In silico* e-probes, analogous to a physical probe or primer, were designed with high specificity to a target organism (a virus, bacterium, mollicute, fungus, or oomycete; Stobbe et al., 2013, Stobbe et al., 2014, Espindola et al., 2015, and Blagden et al., 2016) using a modified version of the Tool for Oligonucleotide Fingerprint Identification (TOFI). The modified TOFI pipeline included the initial step of comparing target sequences (full genome, shotgun, or EST sequences) to near neighbor sequences to identify unique, target-specific sequences. The next step of TOFI is to optimize thermodynamic properties of the identified probes. As thermodynamic optimization is unnecessary for e-probes, this step was replaced with specific BLAST parameters, restriction of e-probe length, and removal of homo-oligomers (five or more consecutive identical nucleotides), as many NGS platforms have problems recognizing these constructs. The TOFI then performs a BLAST search to ensure uniqueness of the probe; however, this step was also modified by Stobbe et al. (2013) to optimize e-probe length then perform a BLAST search. E-probes were then manually edited, if necessary, and eliminated if they matched a non-target sequence at an E-value of 10^{-10} or less. These probes are then used to query the sequence dataset using BLASTn to gauge presence or absence of the target. This assay allows for minimal bioinformatic processing of the large dataset by detecting the presence of sequences of interest while eliminating post-sequencing short read assembly, quality checks, and lengthy annotation.

To date, EDNA has been published as a theoretical method for diagnostic detection in NGS datasets; that is, the approach has been developed for certain pathogens and tested *in silico* using generated mock sample databases (MSDs) (Stobbe et al., 2013). EDNA also has been

validated and published for detection of plant viruses (Stobbe et al., 2014) and fungi and oomycetes (Espindola et al., 2015) within plant-derived NGS datasets and for human pathogens on plants (Blagden et al., 2016). Validation of EDNA for bacteria and mollicute detection in plant NGS datasets is currently awaiting submission for publication, as well (Daniels et al., in preparation). MSDs serve as simulations of NGS runs and can be generated for different NGS platforms (454-pyrosequencing, Illumina, etc.). These include known plant host and pathogen genome sequences present in the sample at different abundances ranging from high pathogen prevalence (15-25% pathogen reads) to very low pathogen prevalence (<0.5%), and multiple MSDs can be designed for each category to serve as replicates. After e-probe design and MSD construction, BLASTn is used to query the MSDs with the pathogen-specific e-probes and a decoy e-probe set that is used for statistical analysis. Decoy e-probe sets are simply the reverse sequences of target-specific e-probes, and their hits in the database represent background signal. To make a diagnostic call (whether the MSD is positive or negative for the target pathogen), the matches of the target-specific e-probes are compared to the matches resulting from the decoy probes, and each e-probe is given a score derived from Equation 1 below (Stobbe et al., 2013). The scores for both sets are then compared using a T-test, and the following calls are made based on the resulting p-value: positive (p-value ≤ 0.05), suspect (p-value ≤ 0.1), or negative (p-value > 0.1). Scores and resulting calls are dependent on the designated BLASTn E-values and the percent coverage of the top e-probe hits.

$$\sum_{h-1}^{n} \{(\%cov.) * [-\log Eval]\}.$$

Equation 1. Equation used to score individual e-probes, where *n* is the top *n* hits equaling [50, 10, 5, 1], *%cov* is the percentage of the e-probe giving the high scoring segment pairing, and *Eval* is the E-value of the *n*th hit (Stobbe et al., 2013).

An alternative method that has been used for calling presence or absence of the target pathogen is the chi-square test. Using this statistic, the matches of the target-specific e-probes and the matches resulting from the decoy probes are used to calculate a chi-square value, which also takes into account the total number of e-probes, to test the goodness of fit between the observed match value and the theoretically expected value. The chi-square value is then converted to a p-value (http://www.socscistatistics.com/pvalues/chidistribution.aspx), and the results are determined to be either statistically significant ($p \leq 0.05$; target is detected or it is not detected) or not statistically significant ($p > 0.05$; detection cannot be determined).

Thus far, e-probes have been designed and tested for nine plant pathogens: *Bean golden mosaic virus*, *Plum pox virus*, *Spiroplasma citri*, *Candidiatus* Liberibacter asiaticus, *Xanthomonas oryzae* pv. *oryzae*, *Xyllela fastidiosa* 9a5c, *Ralstonia solanacearum* race 3 biovar 2, *Puccinia graminis*, *Pythium ultimum*, *Phakopsora pachyrhizi*, *Phytophthora ramorum*, and *E. coli* 0157:H7 (Stobbe et al., 2013; Stobbe et al., 2014; Espindola et al., 2015; Blagden et al., 2016). Published abstracts have included work on EDNA adaptation to general virus discovery in NGS datasets (Stobbe et al., 2013; Dutta et al., in preparation). The adaptability of this method makes it ideal for use on any class of pathogen or microorganism in general and in different datasets. To date, NCBI sequence read archive (SRA) datasets deriving from plant and sewage samples and plant derived NGS datasets, both MSD and actual, have been tested using the EDNA pipeline. This leaves a significant reservoir of plant pathogens and potential candidate for diagnostic metagenomic screening yet untested: insect vectors.

The majority of plant viruses can be transmitted into host plants by insect vectors. Of the approximately 700 known plant virus species, over 75% are transmitted by insects or other mobile vectors (Hogenhout et al., 2008). All phytopathogenic spiroplasmas and phytoplasmas are

transmitted by mobile insect vectors as well, with the exception of the rare case of loss of transmissibility after repeated grafting (Wayadande et al., 1993). With the high rates of insect transmission in these pathosystems, monitoring and avoidance of the majority of viral or mollicute infections could be accomplished via the insect reservoir rather than by the host or pathogen itself. By monitoring known or potential vectors, personnel at ports of entry could know when certain viruses or mollicutes are being carried and can assess when and what interventions should be made. Adapting NGS and EDNA methodologies to this purpose has the potential to be extremely useful, particularly for quick assessment of materials at ports of entry.

Leafhopper Transmission of Plant Pathogens

Viruses

The International Committee on Taxonomy of Viruses (ICTV) classifies approximately 1100 plant-infecting virus species (Hull, 2013). These include primarily ssRNA viruses, some ssDNA viruses represented mostly by the Geminviridae transmitted by whiteflies (Hemiptera; Aleyrodidae), and relatively few enveloped RNA viruses (Table II.1; Hogenhout et al., 2008). Of the viruses having no identified vectors, some are hypothesized to have a vector. The most economically important vectors are classified in the orders Hemiptera and Thysanoptera, the latter of which is a close relative of Hemiptera. These orders have specialized mouthparts that allow for the direct penetration of plant vascular cells and subsequent virus inoculation into host cells. Specifically, hemipterans have piercing-sucking mouthparts composed of a stylet bundle including two mandibular (outer) and two maxillary (inner) stylets (Cranston *et al*., 2009). Two canals are formed by the maxillary stylets; the food canal takes sap up into the vector for nutrition, while the smaller salivary canal releases saliva into the plant while locating a feeding

site. This feeding behavior allows for both the uptake of viruses from plant cells and the inoculation of cells via viruliferous saliva (Storey, 1939).

Table II.1. Summary of insect-transmitted plant viruses (Hogenhout et al., 2008).

Vector taxa	Vector group	Virus groups				Total	%
		Icosahedral particles RNA genome	Rod-shaped particles RNA genome	DNA genome	Enveloped particles RNA genome		
Hemiptera	Aphids	26	153[a]	13	5	197	28
	Whiteflies	–	13	115[b]	–	128	18
	Leafhoppers	8	–	15	3	26	4
	Planthoppers	10	4[c]	–	4	18	3
	Other hemiptera	–	8	5	–	13	2
Thysanoptera	Thrips	2	–	–	14	16	2
Coleoptera	Beetles	50	1	–	–	51	7
Acari	Mites	10	9	–	–	10	1
Nematoda	Nematodes	45	3	–	–	48	7
Mycota	Fungi	8	16	–	–	24	3
	No identified vectors	84	60	19	3[d]	166	24
	Total	233	268	167	30	697	
	%	33	39	24			

[a]Includes 110 virus species of the genus *Potyvirus*, family *Potyviridae*; [b]Virus species of the genus *Begomovirus*, family *Geminiviridae*; [c]These are all tenuiviruses that have multiple shapes; [d]These viruses probably have insect vectors.

Virus transmission by insects was described initially as either non-persistent or persistent based on the duration of retention and transmission (Watson and Roberts, 1939). In non-persistent transmission, inoculation occurred for only minutes after acquisition, while in persistent transmission, virions could be transmitted for days. As previously stated, there are currently four described interactions between plant viruses and an insect vector's body; these include non-

persistent transmission, semi-persistent transmission, persistent-circulative transmission, and persistent-propagative transmission (Watson and Roberts, 1939; Kennedy *et al*., 1962; Sylvester, 1954; Nault, 1997). Each described transmission mode is now based on primarily the length of time in which the virion is retained in the vector and, additionally, the retention site of the virion in the vector (Nault, 1997), although the specific retention site has been shown to vary (Uzest *et al*., 2007). Each plant virus has a very specific, evolved molecular interaction with its insect vector, which has allowed for the emergence of new transmissible variants, as well as a constraint on vast viral variability (Power, 2000). Although each virus-vector interaction is highly specialized, the four modes of transmission generally describe each insect-transmitted virus group and can help to elucidate the movement of newly discovered plant viruses.

As shown in Table II.1, leafhoppers are currently known to transmit 26 plant viruses. This number accounts for a low percentage of known or hypothesized vector-transmitted plant viruses (4%), but leafhoppers remain an important vector group. All leafhopper transmitted pathogens were believed to be viruses until the late 1960s when mulberry dwarf disease, previously thought to be of viral etiology, was found to be caused by mycoplasma-like organisms (MLOs), now known as phytoplasmas (Doi et al., 1967). Soon after, a second group of plant-infecting wall-less bacteria (mollicutes), helical in shape and cultivable, unlike phytoplasmas, was recognized and termed spiroplasmas (Davis et al., 1973). Phytopathogenic phytoplasmas and spiroplasmas have since been shown to be transmitted specifically by insects in the order Hemiptera, suborder Auchenorrhyncha, with most transmitted by leafhoppers (Cicadellidae) and relatively few transmitted by planthoppers (Fulgoromorpha) and psyllids (Psyllidae) (Weintraub and Beanland, 2006). Like viruses, plant mollicutes have specific molecular interactions with their Hemipteran vectors and encounter the same barriers to transmission. These bacteria are transmitted propagatively and, when acquired, remain in the vector for its lifespan (Nault, 1980).

Leafhoppers also are known to transmit walled bacteria, such as *Xylella fastidiosa*, causal agent of Pierce's disease of grapevine (Hewitt et al., 1942).

Phytopathogenic Spiroplasmas

Phytopathogenic spiroplasmas are specialized, plant-infecting, helical, cell wall-less bacteria (Cole et al., 1973). A wide variety of Mollicutes species infecting or inhabiting a number of animal species exist; however, a relatively small number of species are known to infect plants, *S. citri*, *S. kunkelii*, and *S phoeniceum*. Phytopathogenic species are phloem-restricted and transmitted by insect vectors, including primarily leafhoppers. Plant-infecting spiroplasmas are widely thought to have very specific interactions with their insect vectors. After being ingested by a leafhopper from an infected host plant, spiroplasmas are invaginated into the midgut epithelium via membrane-bound vesicles, exit the basal plasmalemma, and cross the basal lamina into the hemolymph (Kwon et al., 1999). In the hemolymph, spiroplasmas avoid vector-related immune degradation, can move into different organs, such as the brain and Malphigian tubules, and eventually enter into the salivary glands using endocytotic processes similar to those exploited in the gut epithelial cells (Granados 1969; Liu et al., 1983). Upon invading the salivary glands, spiroplasmas can be deposited into host plant phloem via the vector's salivary secretions. Adding to the complexity of this plant pathogen-vector interaction, spiroplasmas also multiply while traversing vector tissues and hemolymph. What little is understood about spiroplasma movement through leafhopper vectors to date suggests highly intricate relationships between these pathogens and their vectors that warrant further study.

Spiroplasma-Vector Molecular Interactions

Specific molecular interactions between spiroplasmas and their vectors are largely unknown. Studies on spiroplasma membrane-associated proteins, such as P89, spiralin, and Sc76, have elucidated some interactions between the pathogen and vector cells, but mostly from the perspective of the pathogen. For example, spiroplasma membrane protein P89 (SARP1; spiroplasma adhesion related protein) was discovered to play an important role in attachment of *S. citri* to vector *Circulifer tenellus* cells (Yu et al., 2000; Berg et al., 2001). Using monolayers of *C. tenellus* cell cultures, Yu et al. found that adherence of *S. citri* to vector cells was significantly reduced by treatment with proteinase K or pronase. Restoration of the P89 protein then reestablished *S. citri* cell adherence. *Spiroplasma citri* lipoprotein spiralin, which plays a role in attachment to insect cells, was required for efficient insect transmission in studies using spiralin mutants and spiroplasma overlay assays of protein blots (Duret et al., 2003; Killiny et al., 2005). In the overlay assay, spiralin showed particular affinity for insect-associated 50 kDa to 60 kDa glycoproteins *in vitro*, suggesting potential specific interactions between this lipoprotein and glycoproteins at the insect vector cellular surface; however, the specific identities of these particular insect-associated glycoproteins were not discovered. *S. citri* protein Sc76, hypothesized to be a transmembrane solute binding protein (SBP) of a sugar ABC transporter, is also thought to play a role in insect transmission. G76 (Sc76) mutants had reduced titers in vector salivary glands, and transmission tests showed less efficient transmission and delayed symptom development in the plant host (Boutareaud et al., 2004). Again, particular insect molecules or moieties that this protein potentially interacts with are yet unknown. Other studies with *S. citri* and leafhopper vector *C. haematoceps* have described additional *S. citri* membrane-associated proteins, including P32 and ScARPs (*S. citri* adhesion-related proteins), as potentially playing roles in insect transmission (Berho et al., 2006; Berho et al., 2006; Killiny et al., 2006).

Though most studies to date have focused on the membrane and membrane-associated proteins of *S. citri*, one particular vector-associated protein has been implicated in insect vector specificity. Confocal analysis of *S. citri* internalized in *C. haematoceps* salivary gland cells revealed that the mollicutes locate intracellularly along actin microfilaments (Labroussaa et al., 2010). This interaction with the actin cytoskeleton was also observed in cells of Ciha-1, a *C. haematoceps* embryonic cell line, as well as in other intracellular mollicute (*Candidatus* Phytoplasma asteris, OY strain) and bacterial systems (Cossart et al., 2003; Gouin et al., 2005; Suzuki et al., 2006). Further study of leafhopper-spiroplasma interactions with particular focus on the vector could reveal important vector-related biomolecules required in the transmission process, which is a key aim of this research. These could in turn lead to new strategies for vector transmission management.

Vector System of Interest

Exitianus exitiosus

The gray lawn leafhopper, *E. exitiosus* (Uhler) (Hemiptera: Cicadellidae), is a commonly occurring agricultural pest throughout the North and Central Americas. This species feeds primarily on Poaceae hosts, including cereals, grasses, and forage crops, with susceptible cultivars of *Avena sativa* (oats), *Hordeum vulgare* (barley), and *Triticum aestivum* (wheat) providing some of the best reproductive hosts (Gustin and Stoner, 1973). *E. exitiosus* was first recorded as an agricultural pest in 1879 when it was found to have damaged wheat crops in North Carolina, South Carolina, and Georgia (Osborn, 1912). Little more was reported on the species until the mid-1960s, when it was collected from corn affected with corn stunt disease and suspected as a vector (Douglas et al., 1966).

Although this species is frequently recorded in agricultural pest collections and general insect surveys within the U.S., little research has been conducted on the gray lawn leafhopper. What has been described for *E. exitiosus*, however, is its life history within the laboratory setting (Gustin and Stoner, 1973) and some basic plant pathogen transmission biology (Nault, 1980; Nault and Madden, 1988). In the laboratory setting, the life cycle of *E. exitiosus* from egg to adult emergence is approximately 55 and 60 days for males and females respectively with (typically) 5 nymphal stages (stadia). Females lay up to 50 eggs in groups of 3-6 under the epidermal layer of the leaf sheath over their 17 – 43 (average 32.5) days as adults. This general life cycle data was helpful for determining collection time points and latent period allowances for experimentation in the work described here.

E. exitiosus is an experimental vector of both corn stunt spiroplasma (*S. kunkelii*) and *Maize chlorotic dwarf virus* (MCDV). MCDV is transmitted semi-persistently by its vectors (Nault et al., 1973). Its transmission efficiency by *E. exitiosus*, however, is low. Out of 325 individuals tested, only 12.6% transmitted MCDV (Nault and Madden, 1988). This species is therefore not considered to play a big role in the epidemiology of MCDV in corn. By comparison, the corn stunt spiroplasma is transmitted much more efficiently by the gray lawn leafhopper. Transmission tests for this vector system showed a 78.6% transmission efficiency rate after a 4-day AAP from infected corn, 17-day LP, and 7-day IAP on corn (Nault, 1980). However, only 28 individuals were tested, therefore the efficiency might differ given more tests or under different environmental conditions.

Spiroplasma kunkelii

The corn stunt spiroplasma, *S. kunkelii*, is a cell wall-less, motile, helical bacterium classified in class Mollicutes, order Entomoplasmatales, family Spiroplasmataceae. The majority of spiroplasma species are considered endosymbionts or pathogens of insects, typically found

inhabiting the gut and hemolymph. Few species, including *S. citri*, *S. kunkelii*, and *S. phoeniceum*, are phytopathogenic, infecting plant phloem tissue. Phytopathogenic spiroplasmas cause stunting and discoloration of their plant hosts. *S. kunkelii* infects primarily monocotyledonous *Zea* species, including *Z. mays*, *Z. mays Mexicana*, *Z. diploperennis*, *Z. perennis*, *Z. luxurians*, and *Z. mays* X *Tripsacum floridanum* (Nault, 1980). *Vicia faba* (broad bean), *Catheranthus roseus* (rosy periwinkle), and *Lolium perenne* (perennial ryegrass), *Raphanus sativus* (radish), *Sinapis alba* (mustard), and *Spinacia oleracea* (spinach) are also reported as hosts of *S. kunkelii* (Markham et al., 1977).

Like other phytopathogenic mollicutes, *S. kunkelii* is transmitted by insect vectors, primarily leafhoppers. This species is transmitted to plants naturally by two leafhopper species, *Dalbulus maidis* (DeLong and Wolcott) and *D. elimatus* (Ball), and experimentally by *E. exitiosus* among a few other Deltocephalinae leafhopper species (Granados et al., 1968; Granados, 1969; Nault, 1980). The designation of *E. exitiosus* as an experimental rather than natural vector of *S. kunkelii* is because transmission has been demonstrated only in a laboratory setting. The high experimental transmission efficiency of *S. kunkelii* by *E. exitiosus* that has been observed in the laboratory, however, could reflect a naturally evolved vector system combination.

S. kunkelii is transmitted propagatively by its insect vectors, including *E. exitiosus* (Nault, 1980). Acquisition and transmission are dose-dependent. In *D. maidis*, as the AAP and IAPs increase, transmission efficiencies likewise increase and can reach near 100% (Alivizatos and Markham, 1986). Though *D. maidis* can transmit *S. kunkelii* with very high efficiency given optimal conditions, transmission by other leafhopper species is thought to be much less efficient. A study on the survival and fecundity of leahopper vectors infected with *S. kunkelii* showed that *D. maidis* fitness is not affected by *S. kunkelii*, but other species' survival and fecundity are

significantly reduced (Madden et al., 1984). This difference in effect of the pathogen on *D. maidis* versus other vectors suggests a coevolution between these two species.

Like other bacteria, the *S. kunkelii* genome consists of two components: the chromosomal genome and the accessory genome (Hacker and Kaper, 2000). The chromosome is composed primarily of core metabolism genes and is highly conserved, while the accessory genome contains more pathogenicity-, antibiotic resistance-, and fitness-related genes (Hacker and Carniel, 2001). Studies on spiroplasmal genomes have revealed tendencies for high changes and variations among strains. Genome rearrangements by insertions, deletions, and inversions in *S. citri* are extensive. Chromosomal inversion and 5-10 kb and 10-20 kb deletions at the inverted segment borders were identified in the insect non-transmissible strain BR3-G when compared with two other insect-transmissible strains (Fletcher et al., 1998). Both *S. citri* and *S. kunkelii* genomes have also been altered by DNA of viral origin, including SpV1 plectroviruses (Melcher and Fletcher, 1999; Melcher et al., 1999; Sha et al., 2000; Bai and Hogenhout, 2002). The complete 1.46 Mb *S. kunkelii* CR2-3X strain chromosomal genome and three plasmid sequences were recently made available (Davis et al., 2015).

Justification

The first objective of this research is to design and validate PCR primer sets that are specific to economically significant *B. tabaci* biotypes. The use of biotype-specific PCR primers for the identification of whitefly samples will be beneficial to both researchers of *B. tabaci* and related whiteflies and to inspectors at ports of entry. Further, primers with 5' A/T-rich flaps have been shown to increase PCR sensitivity and yield in end-point PCR reactions (Afonina, 2007). To increase the limit of detection by these primers and the subsequent PCR yield, these primers will

be synthesized with 5’ A/T-rich flaps and products of these PCR reactions will be compared with those resulting from primers lacking 5’ flaps. Other rapid identification methods such as multiplex PCR, helicase dependent amplification (HDA), and melting temperature analysis protocols will also be developed. These approaches will reduce the number of reactions necessary for the identification of certain biotypes, provide an isothermal reaction option, and allow for rapid whitefly biotype discrimination without the need for gel electrophoresis, respectively.

As travel and the trade of agricultural goods among countries increase, the risk of inadvertent introduction of pests and pathogens increases as well. Next generation sequencing provides a significant opportunity for screening any microbes that may enter into the United States in or on agricultural products. EDNA allows for the detection of all pathogen classes in or on plants. As exotic insect vectors can be a significant source of foreign plant pathogens, the second objective of this work is to validate the use of EDNA for detection of plant pathogens within insect vector NGS datasets. As *S. kunkelii* multiplies within its experimental insect vector, *E. exitiosus*, this system will be used as a model for initial testing of EDNA detection within insects versus plants. Validation of this type of vector system can then lead to sensitivity testing and validation within other vector systems, such as *Soybean dwarf virus* (SbDV) transmission by the pea aphid or *Plum pox virus* (PPV) transmission by the green peach aphid. SbDV represents a persistent-circulative virus that lacks propagation within the vector host. PPV represents a non-persistent virus system, which would be an important test of EDNA’s detection limitations.

Further, there is clear evidence for intimate molecular interactions between phytopathogenic mollicutes and insects in the assessment of vector transmission. Plant-infecting mollicutes can have effects on their insect hosts; that is, their invasion into leafhopper tissues can result in either reduced or increased vector fitness depending on the vector and spiroplasma species (Granados and Meehan, 1975; Madden et al., 1984; Bressan et al., 2005; D’Amelio et al.,

2008). I hypothesize that there will be transcriptional responses in *E. exitiosus* to invasion by *S. kunkelii*. Specifically, as *S. kunkelii* traverses and invades *E. exitiosus* cells, there will be up- or down-regulation of particular leafhopper vector genes. The third objective of this study is to discover genes in *E. exitiosus* that are differentially transcribed in response to *S. kunkelii* infection. Total RNA will be extracted from both spiroplasma-exposed/acquired and spiroplasma-naïve *E. exitiosus* and sequenced to determine relative transcript levels. Any substantial differences in gene transcription levels will presumably be due to *S. kunkelii* presence as both groups will be maintained under identical conditions during the duration of the experiment. The genes discovered to be differentially transcribed in response to *S. kunkelii* presence could indicate genes and subsequently proteins that play a role in spiroplasma interaction with the vector and, potentially, vector transmission.

REFERENCES

Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., and Boonham, N. 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Molecular Plant Pathology 10(4):537-45.

Afonina, I., Ankoudinova. I., Mills, A., Lokhov, S., Huynh, P., and Mahoney, W. 2007. Primers with 5′ flaps improve real-time PCR. BioTechniques 43:770-774.

Agranovsky, A.A. 1996. Principles of molecular organization, expression, and evolution of closteroviruses: over the barriers. Advances in Virus Research 47:119-158.

Alivizatos, A. S. and Markham, P. G. 1986. Acquisition and transmission of corn stunt spiroplasma by its leafhopper vector *Dalbulus maidis*. Annals of Applied Biology 108:535-544.

Amari, K., Gonzalez-Ibeas, D., Gómez, P., Sempere, R.N., Sanchez-Pina, M.A., Aranda, M.A., Diaz-Pendon, J.A., Navas-Castillo, J., Moriones, E., Blanca, J. and Hernandez-Gallardo, M.D. 2008. *Tomato torrado virus* is transmitted by *Bemisia tabaci* and infects pepper and eggplant in addition to tomato. Plant Disease 92(7):1139-1139.

Antignus, Y., Perlsman, M., Ben-Yoseph, R., Cohen, S. 1993. The interaction of *Tomato yellow leaf curl virus* with its whitefly vector, *Bemisia tabaci*. Phytoparasitica 21:174-175.

Azzam, O., Frazer, J., Delarosa, D., Beaver, J.S., Ahlquist, R, and Maxwell, D. P. 1994.Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus require functional coat protein. Virology 204(1):289-96.

Bai, X. and Hogenhout, S.A. 2002. A genome sequence survey of the mollicute corn stunt spiroplasma *Spiroplasma kunkelii*. FEMS microbiology letters 210(1):7-17.

Basu, A.N. 1995. *Bemisia tabaci* (Gennadius) Crop pest and principal whitefly vector of plant viruses. Westview press: Boulder. 180 pp.

Berg, M., Melcher, U. and Fletcher, J. 2001. Characterization of Spiroplasma citri adhesion related protein SARP1, which contains a domain of a novel family designated sarpin. Gene 275(1):57-64.

Berlinger, M.J. 1986. Host Plant Resistance to *Bemisia tabaci*. Agriculture, Ecosystems, & Environment 17:69-82.

Baumann, P., Munson, M. A., Lai, C.Y., Clark, M.A., Baumann, L., Moran, N.A. and Campbell, B.C. 1993. Origin and properties of bacterial endosymbionts of aphids, whiteflies, and mealybugs. American Society for Microbiology News 59(1):21-24.

Bedford, I.D., Briddon, R.W., Brown, J.K., Rosell, R.C., Markham, P.G. 1994. Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Gennadius) biotypes from different geographic region. Annals of Applied Biology 125(2):311-325

Bellows, T. S., Perring, T. M., Gill, R. J., & Headrick, D. H. 1994. Description of a species of *Bemisia* (Homoptera: Aleyrodidae). Annals of the Entomological Society of America 87(2):195-206.

Berho N, Duret S, Danet JL, and Renaudin J. 2006. Plasmid pSci6 from Spiroplasma citri GII-3 confers insect transmissibility to the non-transmissible strain *S. citri* 44. Microbiology 152: 2703–2716.

Berho N, Duret S, and Renaudin J. 2006. Absence of plasmids encoding adhesion-related proteins in non-insect-transmissible strains of *Spiroplasma citri*. Microbiology 152: 873–886.

Bird, J. 1957. A Whitefly-transmitted Mosaic of *Jatropha gossypifolia*. Technical Papers. Porto Rico Agricultural Experiment Station, Insular Station, Rio Piedras. 35 pp.

Bird, J., and Maramorosch, K. 1978. Viruses and virus diseases associated with whiteflies. Advancements in Virus Research 22:55-110.

Blagden, T., Schneider, W., Melcher, U., and Fletcher, J. 2013. *In silico* Adaptation of EDNA (E-probe Diagnostic Nucleic Acid Analysis) for Detection of Foodborne Pathogens. Annual Meeting, July 28-31, 2013, IAFP.

Bosco, D., Mason, G., and Accotto, G.P. 2004. TYLCSV DNA, but not infectivity, can be transovarially inherited by the progeny of the whitefly vector *Bemisia tabaci* (Gennadius). Virology 323(2):276-283.

Boutareaud, A., Danet, J. L., Garnier, M., and Saillard, C. 2004. Disruption of a gene predicted to encode a solute binding protein of an ABC transporter reduces transmission of Spiroplasma citri by the leafhopper Circulifer haematoceps. Applied and environmental microbiology 70:3960-3967.

Bradfute, O.E., Tsai, J.H., and Gordon, D.T. 1981. Corn stunt spiroplasma and viruses associated with a maize disease epidemic in Southern Florida. Plant Disease 65(10):837-841.

Bressan, A., Girolami, V., and Boudon-Padieu, E. 2005. Reduced fitness of the leafhopper vector Scaphoideus titanus exposed to Flavescence dorée phytoplasma. Entomologia experimentalis et applicata 115:283-290.

Briddon, R.W., Pinner, M.S., Stanley, J., and Markham, R.G. 1990. Geminivirus coat protein gene replacement alters insect specificity. Virology 177(1):85-94.

Brown, J. K. 1994. The Status of *Bemisia tabaci* Genn. as a pest and vector in world agroecosystems. FAO Plant Protection Bulletin 42:3-32.

Brown, J.K., Frohlich, D.R., and Rosell, R.C. 1995. The Sweetpotato or Silverleaf Whiteflies: Biotypes of *Bemisia tabaci* or a Species Complex? Annual Review of Entomology 40:511-534.

Brown, J.K. 2001. The molecular epidemiology of begomoviruses. In 'Trends in Plant Virology'. J. A. Khan and J. Dykstra, ed. The Haworth Press Inc., New York.

Brown, J., and Czosnek. 2002. Whitefly transmission of plant viruses. Advances in Botanical Research 36: 65-76.

Buchner, P. 1965. Endosymbiosis of animals with plant microorganisms. Interscience Publishers, New York.

Butler, G.D., Henneberry, T.J., Clayton, T.E. 1983. *Bemisia tabaci* (Homoptera:Aleyrodidae): development, oviposition, and longevity in relation to temperature. Annals of the Entomological Society of America 76:310-13.

Byrne, F.J. and Devonshire, A.L. 1996. Biochemical evidence of haplodiploidy in the whitefly *Bemisia tabaci*. Genetics 34:93-107.

Capoor, S. E. 1949. Feeding methods of the whitefly. Current Science 19:248-249.

Carpane, P., Melcher, U., Wayadande, A., de la Paz Gimenez Pecci, M., Laguna, G., Dolezal, W., and Fletcher, J. 2013. An analysis of the genomic variability of the phytopathogenic mollicute Spiroplasma kunkelii. Phytopathology 103(2):129-134.

Carus, W.S. 2000. Toxic Terror: Assessing Terrorist Use of Chemical and Biological Weapons. Cambridge, Massachusetts: MIT Press.

Cervera, M.T., Cabezas, J.A., Simón, B., Martínez-Zapater, J.M., Beitia, F., and Cenis, J.L. 2000. Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on AFLP analysis. Bulletin of Entomological Research 90:391-396.

Chalk, P. 2001. Testimony on terrorism, infrastructure protection, and the U.S. food and agricultural sector. Hearings Before the Senate Subcommittee on Oversight of

Government Management, Restructuring, and the District of Columbia, 107th Cong, 1st Sess.

Cicero, J.M. and Brown, J.K. 2011a. Functional Anatomy of Whitefly Organs Associated with Squash Leaf Curl Virus (Geminiviridae: Begomovirus) Transmission by the B Biotype of Bemisia tabaci (Hemiptera: Aleyrodidae). Annals of the Entomological Society of America 104(2):261-279.

Cicero, J.M. and Brown, J.K. 2011b. Anatomy of Accessory Salivary Glands of the Whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) and Correlations to Begomovirus Transmission. Annals of the Entomological Society of America 104(2):280-286.

Cieslak, T.J. and Eitzen, E.M. 1999. Clinical and epidemiologic principles of anthrax. Emerging Infectious Diseases 5(4):552–555.

Clark, M.A., Baumann, L., Munson, M.A., Baumann, P., Campbell, B.C., Duffus, J.E., Osborne, S. L., and Moran, N.A. 1992. The eubacterial endosymbionts of whiteflies (Homoptera: Aleyrodidae) constitute a lineage distinct from the endosymbionts of aphids and mealybugs. Current Microbiology 25(2):119-123.

Cock, M.J.W. 1986. *Bemisia tabaci* - a literature survey on the cotton whitefly with an annotated bibliography. 121 pp.

Cohen, A.C., Henneberry, T.J., and Chu, C.C. 1996. Geometric relationships between whitefly feeding behavior and vascular bundle arrangements. Entomologia Experimentalis et Applicata 78:135-142.

Cohen, A.C., Chu, C.C., Henneberry, T.J., Freeman, T., Nelson, D., Buckner, J., Margosan, D.,Vail, P., and Aung, L.H. 1998. Feeding biology of the silverleaf whitefly (Homoptera: Aleyrodidae). Chinese Journal of Entomology 18:65-81.

Cohen, S., and Nitzany, F.E. 1966. Transmission and host range of the *Tomato yellow leaf curl virus*. Phytopathology 56:1127–1131.

Cole, R.M., Tully, J.G., Popkin, T.J., and Bové, J.M. 1973. Morphology, ultrastructure, and bacteriophage infection of the helical mycoplasma-like organism (Spiroplasma citri gen. nov., sp. nov.) cultured from “stubborn” disease of citrus. Journal of bacteriology 115(1):367-386.

Cossart, P., Pizarro-Cerda, J., and Lecuit, M. 2003. Invasion of mammalian cells by Listeria monocytogenes: functional mimicry to subvert cellular functions. Trends in Cell Biology 13: 23–31.

Costa, A.S. 1969. Whiteflies as virus vectors. In 'Viruses, Vectors, and Vegetation'. K. Maramorosch, ed. Interscience Press, New York pp. 95-119.

Costa, H.S., and Brown, J.K. 1991. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction. Entomologia Experimentalis et Applicata. 61:211–19.

Costa, H.S., Brown, J.K., Sivasupramaniam, S. and Bird, J., 1993. Regional distribution, insecticide resistance, and reciprocal crosses between the A and B biotypes of Bemisia tabaci. International Journal of Tropical Insect Science 14(02):255-266.

Costa, H.S., Westcot, D.M., Ullman, D.E., Rosell, R., Brown, J.K., and Johnson, M.W. 1995. Morphological variation in Bemisia endosymbionts. Protoplasma 189:194-202.

Czosnek, H., Ghanim, M., Ghanim, M. 2002. The circulative pathway of begomoviruses in the whitefly vector *Bemisia tabaci*— insights from studies with *Tomato yellow leaf curl virus.* Annals of Applied Biology 140(3):215–231.

D'Amelio, R., Palermo, S., Marzachi, C., and Bosco, D. 2008. Influence of Chrysanthemum yellows phytoplasma on the fitness of two of its leafhopper vectors, Macrosteles quadripunctulatus and Euscelidius variegatus. Bulletin of Insectology 61: 349-354.

Damsteegt, V.D., Stone, A.L., Kuhlmann, M., Gildow, F.E., Domier, L.L., Sherman, D.J., Tian, B., and Schneider, W.L. 2011. Acquisition and Transmissibility of US Soybean dwarf virus Isolates by the Soybean Aphid, *Aphis glycines.* Plant Disease 95: 945-950.

Daniels, J., Schneider, B., and Ochoa-Corona, F. Detection of prokaryotic plant pathogens using next generation sequencing and electronic probes (EDNA). In preparation for submission to BMC Microbiology.

Davis, R. E. and Worley, J. F. 1973. Spiroplasma: motile, helical microorganism associated with corn stunt disease. Phytopathology.

Davis, R.E., Shao, J., Dally, E.L., Zhao, Y., Gasparich, G.E., Gaynor, B.J., Athey, J.C., Harrison, N.A., and Donofrio, N. 2015. Complete genome sequence of *Spiroplasma kunkelii* strain CR2-3x, causal agent of corn stunt disease in *Zea mays* L. Genome Announcements 3(5):e01216-15.

De Barro, P.J. 1995. *Bemisia tabaci* biotype B; A Review of its Biology, Distribution and Control. CSIRO Division of Entomology. Technical Paper No. 36. Canberra, Australia.

De Barro, P.J., Liu, S., Boykin, L.M., and Dinsdale, A.B. 2010. *Bemisia tabaci*: A statement of species status. Annual Review of Entomology 56:1-19.

Dennehy, T. J., Degain, B. A., Harpold, V. S., Brown, J. K., Morin, S., Fabrick, J. A., Byrne, F. J., and Nichols, R. L.. 2005. New challenges to management of whitefly resistance to insecticides in Arizona. The University of Arizona-Cooperative Extension Report, Tucson, AZ.

Doi, Y., Teranaka, M., Yora, K., and Asuyama, H. 1967. Mycoplasma-or PLT group-like microorganisms found in the phloem elements of plants infected with mulberry dwarf, potato witches' broom, aster yellows, or paulownia witches' broom. Annals of Phytopathological Society of Japan. 33:259-266.

Douglas, W. A., Whitcomb, W. H., Hepner, L. W., Kirk, V. M., and Davis, R. 1966. Some Cicadellidae (Homoptera) collected from corn in the southeastern United States. Annals of the Entomological Society of America 59:393-6.

Duret, S., Berho, N., Danet, J. L., Garnier, M., and Renaudin, J. 2003. Spiralin is not essential for helicity, motility, or pathogenicity but is required for efficient transmission of Spiroplasma citri by its leafhopper vector *Circulifer haematoceps*. Applied and Environmental Microbiology 69:6225-6234.

Dutta, M., Stobbe, A., Ali, A., Schneider, W., and Melcher, U. Generalization of e-brobe diagnostic nucleic acid analysis to screen metagenomics data for viruses. In preparation for submission.

Elbaz, M., Lahav, N. Morin, S. 2010. Evidence for pre-zygotic reproductive barrier between the B and Q biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). Bulletin of Entomological Research 100(5):581-590.

Espindola, A., Schneider, W., Hoyt, P. R., Marek, S. M., and Garzon, C. 2015. A new approach for detecting fungal and oomycete plant pathogens in next generation sequencing metagenome data utilising electronic probes. International journal of data mining and bioinformatics 12(2):115-128.

Filichkiu, S.A., Brumfield, S., Filichkin, T.E., and Young, M.J. 1997. In vitro interactions of the aphid endosymbiont SymL chaperonin with *Barley yellow dwarf virus*. Journal of Virology 71(1):569-77.

Fisher, D.B., Wright, J. P., and Mittler, T.E. 1984. Osmoregulation by the aphid *Myzus persicae*: a physiological role for honeydew oligosaccharides. Journal of Insect Physiology 30:387–393.

Fletcher, J., Wayadande, A., Melcher, U., and Ye, F. 1998. The phytopathogenic mollicute-insect vector interface: A closer look. Phytopathology 88:1351-58.

Frohlich, D.R., Torres-Jerez, I.I., Bedford, I.D., Markham, P.G., and Brown, J.K. 1999. A phylogeographical analysis of the *bemisia tabaci* species complex based on mitochondrial DNA markers. Molelcular Ecology 8(10):1683-91.

Garrett, A., Kerlan, C., and Thomas, D. 1993. The intestine is a site of passage for potato leafroll virus from the gut lumen into the haemocoel in the aphid vector, *Myzus persicae* Sulz. Archives of Virology 131:377-392.

Gennadius, P. 1889. Disease of tobacco plantations in the Trikonia. The aleurodid of tobacco. Ellenike Georgia 5:1-3.

Ghanim, M., Morin, S., Zeidan, M., Czosnek, H. 1998. Evidence for transovarial transmission of *Tomato yellow leaf curl virus* by its vector, the whitefly *Bemisia tabaci*. Virology 240(2):295-303.

Ghanim, M., Morin, S., and Czosnek, H. 2001. Rate of *Tomato yellow leaf curl virus* (TYLCV-ls) translocation in the circulative transmission pathway in its vector, the whitefly *Bemisia tabaci.* Phytopathology 91(2):188-96.

Goodman, R.M. 1977. Single-stranded DNA genome in a whitefly-transmitted plant virus. Virology 83(1):171-179.

Gottwald, T.R., Dagraca, J.V., and Bassanezi, R.B. 2007. Citrus huanglongbing: The pathogen and its impact. Online. Plant Health Progress doi: 10.1094/PHP-2007-0906-01-RV.

Gouin, E., Welch, M. D., and Cossart, P. 2005. Actin-based motility of intracellular pathogens. Current Opinion in Microbiology 8:35–45.

Granados, R. R. 1969. Maize viruses and vectors. Pp. 327-59 in K. Maramorosch (ed.) Viruses, vectors and vegetation. Interscience Publishers, New York.

Granados, R. R., Granados, J. S., Maramorosch, K., and Reinitz., J. 1968. Corn stunt virus: transmission by three cicadellid vectors. Journal of Economic Entomology 61:1282-87.

Granados, R. R. and Meehan, D. J. 1975. Pathogenicity of the corn stunt agent to an insect vector, Dalbulus elimatus. Journal of Invertebrate Pathology 26:313-320.

Gray, S., and Banerjee, N. 1999. Mechanisms of Arthropod Transmission of Plant and Animal Viruses. Microbiology and Molecular Biology Reviews 63(1):128-148.

Gray, S., and Gildow, F.E. 2003. Luteovirus-aphid interactions. Annual Review of Phytopathology 41:539-66.

Gustin, R. D. and Stoner, W. N. 1973. Life history of *Exitianus exitiosus* (Homoptera: Cicadellidae) in the laboratory. Annals of the Entomological Society of America 66(2):388-389.

Hacker, J. and Kaper, J. 2000. Pathogenicity islands and the evolution of microbes. Annual Review of Microbiology 54:641-79.

Hacker, J. and Carniel, E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. EMBO Reports 2:376-81.

Halbert, S.E. and Manjunath, K.L. 2004. Asian Citrus Psyllids (Sternorrhyncha: Psyllidae) and Greening Disease in Citrus: A Literature Review and Assessment of Risk in Florida. Florida Entomologist 87(3):330-353.

Hanley-Bowdoin, L., Settlage, S.B., Orozco, B.M., Nagar, S., and Robertson, D. 2000. Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. Critical Reviews in Biochemistry and Molecular Biology 35(2):105-40.

Harrison, B.D. and Murant, A.F. 1984. Vectors in Virus Biology. Academic Press: London. 188 pp.

Hewitt, W. B., Frazier, N. W., Jacob, H. E., and Freitag, J. H. 1942. Pierce's disease of grapevines. California Agricultural Experiment Station 353:1-32

Hodges, G.S. and Evans, G.A. 2005. An Identification Guide to the Whiteflies (Hempitera: Aleyrodidae) of the Southeastern United States. Florida Entomologist 88(4):518-534.

Hofer, E., Bedford, I.D., Markham, E.G., Jeske, H., and Frischmuth, T. 1997. Coat protein gene replacement results in whitefly transmission of an insect non-transmissible geminivirus isolate. Virology 236(2):288-95.

Hogenhout, S.A., Ammar, E., Whitfield A.E., Redinbaugh, M.G. 2008. Insect Vector Interactions with Persistently Transmitted Viruses. Annual Review of Phytopathology 2008.46:327-359.

Hollings, M., Stone, O. M., and Bock, K. R.1976. Purification and properties of sweetpotato mild mottle virus, a whitefly-borne vi-rus from sweet potato in East Africa. Annals of Applied Biology 82:511-528.

Honle, M., Hofer, E, Bedford, I.D., Briddon, R.W., Markham, E.G. and Frischmuth, T. 2001. Exchange of Three Amino Acids in the Coat Protein Results in Efficient Whitefly Transmission of a Nontransmissible *Abutilon mosaic virus* Isolate. Virology 290(1):164-171.

Horowitz, A.R., Kontsedalov, S., Khasdan, V. and Ishaaya, I. 2005. Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. Archives of Insect Biochemistry and Physiology 58(4):216-225.

Hull, R., 2013. Plant virology. Academic press.

Hulme, P.E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46:10-18.

Hunter, W.B., Hiebert, E., Webb, S. E., Tsai, J. H., and Polston, J. E. 1998. Location of Geminiviruses in the Whitefly *Bemisia tabaci* (Homoptera: Aleyrodidae). Plant Disease 82(10):1147-1151.

Janssen, J.A.M, Tjallingnii, W.F., van Lenteren, J.C. 1989. Electrical recording and ultrastruccture of stylet penetration by the greenhouse whitefly. Entomologia Experimentalis et Applicata 52:69-81.

Jelkmann, W., Fechtner, B., and Agranovsky, A.A. 1997. Complete genome structure and phylogenetic analysis of little cherry virus, a mealybug-transmissible closterovirus. Journal of General Virology 78:2067-2071.

Jeyanandarajah, P. and Brunt, A.A. 1993. The natural occurrence, transmission, properties and possible affinities of cowpea mild mottle virus. Phytopathology 137:148-156.

Jiang, Y. X., Lei, H., Collar, J. L., Martín, B., Muñiz, M., and Fereres, A. 1999. Probing and feeding behavior of two distinct biotypes of *Bemisia tabaci* (Homoptera: Aleyrodidae) on tomato plants. Journal of Economic Entomology 92:357–366.

Jiang, Y. X., De Bias, C., Barrios, L. and Fereres, A. 2000. Correlation between whitefly (Homoptera: Aleyrodidae) feeding behavior and transmission of *Tomato yellow leaf curl virus*. Entomological Society of America 93:573-579.

Jones, D. R. 2003. Plant viruses transmitted by whiteflies. Euro. J. Plant Pathol. 109:195-219.

Kegler, H., and W. Hartman. 1998. Present status of controlling conventional strains of plum pox virus. In: Hadidi, A., Khetarpal, R.K., and H. Koganezawa, eds. Plant Virus Disease Control. Pp. 616-628. APS Press, St. Paul, MN.

Kennedy, J.S., Day, M.F., Eastop, V.F. 1962. A Conspectus of Aphids as Vectors of Plant Viruses. London: Commonwealth Institute of Entomology 114 pp.

Khasdan, V., Levin, I., Rosner, A., Morin, S., Kontsedalov, S., Maslenin, L., and Horowitz, A.R. 2005. DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera:Aleyrodidae) and studying population dynamics. Bulletin of Entomological Research 95:605–613.

Killiny, N., Castroviejo, M., and Saillard, C. 2005. Spiroplasma citri spiralin acts in vitro as a lectin binding to glycoproteins from its insect vector Circulifer haematoceps. Phytopathology 95: 541-548.

Killiny, N., Batailler, B., Foissac, X., Saillard, C. 2006. Identification of a Spiroplasma citri hydrophilic protein associated with insect transmissibility. Microbiology 152: 1221–1230.

Kwon, M., Wayadande, A., and Fletcher, J. 1999. Spiroplasma citri movement into the intestines and salivary glands of its leafhopper vector, Circulifer tenellus. Phytopathology 89: 1144-1151.

Labroussaa, F., Arricau-Bouvery, N., Dubrana, M. P., and Saillard, C. 2010. Entry of Spiroplasma citri into Circulifer haematoceps Cells Involves Interaction between Spiroplasma Phosphoglycerate Kinase and Leafhopper Actin. Applied Environmental Microbiology 76: 1879–1886.

Lee, M., SungBok, A., and WangSu, C. 2000. Morphological characteristics of Bemisia tabaci (Gennadius) (Homoptera: Aleyrodidae) and discrimination of their biotypes in Korea by DNA markers. Korean Journal of Applied Entomology 39(1):5-12.

Lei, H., Tjallingii, W.F., and van Lenteren, J.C. 1998. Probing and feeding characteristics of the greenhouse whitefly in association with host-plant acceptance and whitefly stains. Entomologia Experimentalis et Applicata 88:73–80.

Liu, H. Y., Gumpf, D. J., Oldfield, G. N., and Calavan, E. C. 1983. The relationship of Spiroplasma citri and Circulifer tenellus. Phytopathology 73: 585-590.

Ma, D., Hadjistylli, M., Gorman, K., Denholm, I., and Devine, G. 2004. Pre- and post-zygotic breeding incompatibilities between B and Q biotypes of *Bemisia tabaci*. Proceeding of the 2nd European Whitefly Symposium. Cavtat, Croatia.

Madden, L. V., Nault, L. R., Heady, S. E., and Styer, W. E. 1984. Effect of maize stunting mollicutes on survival and fecundity of *Dalbulus* leafhopper species. Annals of Applied Biology 105:431-41

Mardis, E.R. 2008. The impact of next-generation sequencing technology on genetics. Trends in genetics 24(3):133-41.

Mardis, Elaine R. 2013. Next-Generation Sequencing Platforms. Annual Review of Analytical Chemistry 6: 287-303.

Markham, P. G., Townsend, R., Plaskitt, K., and Saglio, P. 1977. Transmission of corn stunt to dicotyledonous plants. Plant Disease Reporter 61:342-45.

Martin, J., Bruno, V. M., Fang, Z., Meng, X., Blow, M., Zhang, T., Sherlock, G., Snyder, M., and Wang, Z. 2010. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics 11: 663.

McCullough, D. G., Work, T. T., Cavey, J. F., Liebhold, A. M., & Marshall, D. 2006. Interceptions of nonindigenous plant pests at US ports of entry and border crossings over a 17-year period. Biological Invasions 8(4):611-630.

McLean, D.L. and Kinsey, M.G. 1964. A technique for electronically recording aphid feeding and salivation. Nature 1358–1359.

Melcher, U. and Fletcher, J., 1999. Genetic variation in Spiroplasma citri.*European journal of plant pathology*, *105*(6), pp.519-533.

Melcher, U., Sha, Y., Ye, F., and Fletcher, J. 1999. Mechanisms of spiroplasma genome variation associated with SpV1-like viral DNA inferred from sequence comparisons. Comparative Microbial Genomics 4(1):29-46.

Meyerson, L.A. and Reaser, J.K. 2002. Biosecurity: moving toward a comprehensive approach to biosecurity is necessary to minimize the risk of harm caused by non-native organisms to agriculture, the economy, the environment, and human health. BioScience 52(7):593-600.

Mound, L.A., and Halsey, S.H. 1978. Whiteflies of the World. New York: Wiley. 340 pp.

Mound, L.A. 1983. Biology and identity of whitefly vectors of plant pathogens. Plant Virus Epidemiology. Oxford: Blackwell Scientific pp. 305-13.

Morin, S., Ghanim, M., Zeidan, M., Czosnek, H., Verbeek, M., and van den Heuvel, J.F.J.M. 1999. A GroEL Homologue from Endosymbiotic Bacteria of the Whitefly *Bemisia tabaci* Is Implicated in the Circulative Transmission of Tomato Yellow Leaf Curl Virus. Virology 256(1):75-84.

Morin, S., Ghanim, M., Sobol, I, and Czosnek, H. 2000. The GroEL protein of the whitefly *Bemisia tabaci* interacts with the coat protein of transmissible and non-transmissible begomoviruses in the yeast two-hybrid system. Virology 276(2):404-16.

Muniyappa, V. and Reddy, D.V.R. 1983. Transmission of cowpea mild mottle virus by *Bemisia tabaci* in a nonpersistent manner. Plant Disease 67:391-393.

Nault, L. R. 1980. Maize bushy stunt and corn stunt: a comparison of disease symptoms, pathogen host ranges, and vectors. Phytopathology 70: 659-662.

Nault, L.R. 1997. Arthropod transmission of plant viruses: a new synthesis. Ann. Entomol. Soc. Am. 90:521–41.

Nault, L. R., and Madden, L. V. 1988. Phylogenetic relatedness of maize chlorotic dwarf virus leafhopper vectors. Phytopathology 78: 1683-1687.

Nault, L. R., Styer, W. E., Knoke, J. K., and Pitre, H. N. 1973. Semipersistent transmission of leafhopper-borne *maize chlorotic dwarf virus*. Journal of Economic Entomology, 66(6), 1271-1273.

Osborn, H. 1912. Leafhoppers affecting cereals, grasses, and forage crops. USDA Bureau Entomology Bulletin 108. 123 p.

Pascual, S., Callejas, C. 2004. Intra- and interspecific competition between biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Spain. Bulletin of Entomological Research 94(4):369-75.

Pollard, D.G. 1955. Feeding habits of the cotton whitefly, *Bemisia tabaci* Genn. (Homoptera: Aleyrodidae). Annals of Applied Biology 43(4):664-671.

Polston, J.E., Al-Musa, A., Perring, T.M., Dodds, J.A. 1990. Association of the nucleic acid of *Squash leaf curl geminivirus* with the whitefly *Bemisia tabaci*. Phytopathology 80:850-856.

Purcell, A.H. 1982. Insect vector relationship with procaryotic plant pathogens. Annual Review of Phytopathology 20:397–417

Quaintance, A.L. 1900. Contribution towards a monograph of the American Aleurodidae. US Department of Agriculture, Technical Services, Bureau of Entomology 8:9-64.

Reisman, D., Ricciardi, R.P., Goodman, R.M. The size and topology of single-stranded DNA from bean golden mosaic virus. Virology 97(2): 388-395.

Richter, B.G. and Sexton, D.P. 2009. Managing and analyzing next-generation sequence data. PLoS Computational Biology 5.6 e1000369.

Rosell, R.C., Torres-Jerez, I., and Brown, J.K. 1999. Tracing the geminivirus-whitefly transmission pathway by polymerase chain reaction in whitefly extracts, saliva, hemolymph, and honeydew. Phytopathology 89(3):239-46.

Rubinstein, G., and Czosnek, H. 1997. Long-term association of *Tomato yellow leaf curl virus* (TYLCV) with its whitefly vector *Bemisia tabaci*: effect on the insect transmission capacity, longevity and fecundity. Journal of Genomic Virology 78: 2683–2689.

Russell, L.M. 1957. Synonyms of *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae). Bulletin of Brooklyn Entomology Society 52: 122-23.

Sanderfoot, A.A. and Lazarowitz, S.G. 1996. Getting it together in plant virus movement: cooperative interactions between bipartite geminivirus movement proteins. Trends in Cell Biology 6:353-358.

Sanger F., Nicklen S., Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Science USA 74: 5463–67.

Schloss, P.D. and Handelsman, J. 2005. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome biology 6(8):229.

Sha, Y., Melcher, U., Davis, R.E., and Fletcher, J. 2000. Common elements of spiroplasma plectroviruses revealed by nucleotide sequence of SVTS2. Virus genes 20(1):47-56.

Shendure, J. and Hanlee, J. 2008. Next-generation DNA sequencing. Nature biotechnology 26: 1135-1145.

Silva, R.G., Galvão, J.C.C., Miranda, G.V., and OLIVEIRA, E.D. 2003. Controle genético da resistência aos enfezamentos do milho. Pesquisa Agropecuária Brasileira 38:921-928.

Stobbe, A.H, Daniels, J., Espindola, A.S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J., and Schneider, W. 2013. E-probe diagnostic nucleic acid analysis (EDNA): a theoretical approach for handling of next generation sequencing data for diagnostics. Journal of Microbiological Methods 94:356-366.

Stobbe, A., W. Schneider, and U. Melcher. 2013. A bioinformatic pipeline for use in metagenomic virus discovery. Annual Meeting. American Phytopathological Society. Abstract.

Suzuki, S., Oshima, K., Kakizawa, S., Arashida, R., Jung, H. Y., Yamaji, Y., Nishigawa, H., Ugaki, M., and Namba, S. 2006. Interaction between the membrane protein of a pathogen and insect microfilament complex determines insect-vector specificity. Proceedings of the National Academy of Science USA 103:4252–4257.

Sylvester, E.S. 1958. Aphid transmission of plant viruses. International Congress of Entomology Montreal, Canada. Presentation.

Tamada, T. 1975. Studies on the soybean dwarf disease. Report of Hokkaido Prefect Agricultural Experiment Station 25:114-144.

Tembe, W., Zavaljevski, N., Bode, E., Chase, C., Geyer, J., Wasieloski, L., Benson, G., and Reifman, J. 2007. Oligonucleotide fingerprint identification for microarray-based pathogen diagnostic assays. Bioinformatics 23(1):5-13.

Tjallingii, W.F. 1978. Electronic recording of penetration behavior by aphids. Entomologia Experimentalis et Applicata 24(3):721–730.

van den Heuvel, J.F., Verbeek, M., van der Wilk, F. 1994. Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus by Myzus persicae. Journal of Genomic Virology 75:2559– 2565.

Varma, A. and Malathi, V.G. 2003. Emerging geminivirus problems: A serious threat to crop production. Annals of Applied Biology, 142:145–164.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., and Gocayne, J.D. 2001. The sequence of the human genome. Science 291(5507):1304-1351.

Vincent, M., Xu, Y., and Kong, H. 2004. Helicase-dependent isothermal DNA amplification. EMBO reports 5:795 – 800.

Waage, J.K. and Mumford, J.D. 2008. Agricultural biosecurity. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363(1492):863-876.

Walker, G.P. and Perring, T.M. 1994. Feeding and Oviposition Behavior of Whiteflies (Homoptera: Aleyrodidae) Interpreted from AC Electronic Feeding Monitor Waveforms. Annals of the Entomological Society of America 87(3):363-374.

Walker, G.P. 1997. Feeding behavior of adult *Bemisia argentifolii* on host and non-host plants. USDA-ARS 103.2.

Watson, M.A., and Roberts, F.M. 1939. A comparative study of the transmission of Hyocyamus virus 3, potato virus Y and cucumber virus 1 by the vectors *Myzus persicae* (Sulz.), *M. circumflexus* (Buckton), and *Macrosiphum gu* (Koch). Proceedings of the Royal Society of London 127:543–76.

Wayadande, A. C., Shaw, M. E., Fletcher, J. 1993. Tests of differential transmission of three *Spiroplasma citri* lines by the leafhopper, *Circulifer tenellus*. Phytopathology 83(4):468-469.

Weintraub, P. G. and Beanland, L. 2006. Insect vectors of phytoplasmas. Annual Review of Entomology 51:91-111.

Wintermantel, W.M. 2004. Emergence of Greenhouse Whitefly (*Trialeurodes vaporariorum*) Transmitted Criniviruses as Threats to Vegetable and Fruit Production in North America. APS*net* Features. Online. doi:10.1094/APSnetFeature-2004-0604

Wu, Z.C., Hu, J.S., Polston, J.E., UIlman, D.E. and Hiebert, E. 1996. Complete nucleotide sequence of a nonvector-transmissible strain of *Abutilon mosaic geminivirus* in Hawaii. Phytopathology 86:608-613.

Xu, J., De Barro, P.J., and Liu, S.S. 2010. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. Bulletin of entomological research 100(03):359-366.

Yu, J., Wayadande, A.C., and Fletcher, J. 2000. *Spiroplasma citri* surface protein P89 implicated in adhesion to cells of the vector *Circulifer tenellus*. Phytopathology 90:716–722.

Zouba, A.A. and Lopez, M.V. 1998. Squash Yellow Leaf Curl Virus: A New Whitefly-Transmitted Poty-Like Virus. Plant Disease 82(5):475-478.

CHAPTER III

RELIABLE SINGLE-TARGET AND MULTIPLEX DISCRIMINATION OF HIGH-CONSEQUENCE *BEMISIA TABACI* BIOTYPES AND *TRIALEURODES VAPORARIORUM* WITH FINE TUNING OF MODIFIED OLIGONUCLEOTIDE PRIMER THERMODYNAMICS

Abstract

The whitefly (Hemiptera; Aleyrodidae) species *Bemisia tabaci* (Gennadius) and *Trialeurodes vaporariorum* (Westwood) are worldwide agricultural pests and virus vectors. *B. tabaci* is composed of a sibling species complex with genetically distinct but morphologically indistinguishable biotypes. Quick identification of whitefly vector species and biotypes can facilitate interventions that prevent cross-border introductions of exotic whiteflies or plant viruses. Whitefly type-specific mitochondrial cytochrome oxidase c I PCR primer sets were designed for the exotic B and Q biotypes and the New World A-like biotype group of the sweetpotato whitefly, *B. tabaci*, and for the greenhouse whitefly, *T. vaporariorum*. These primers were also modified by incorporating A/T-rich overhang sequences at the 5' terminus of each primer (5' flap) in an attempt to increase primer sensitivity. Each primer was evaluated for a change in Gibbs free energy (ΔG) resulting from the 5' flap addition. Single-target endpoint PCRs using each of the eight primer sets were carried out on genomic DNA extracted from

individual adults of each target biotype, and PCR products were separated by electrophoresis in a 2% agarose gel. Resulting PCR amplicons for the three *B. tabaci* biotypes and *T. vaporariorum* primer sets were 712-, 550-, 329-, and 252-bp, respectively, using primer sets lacking overhang sequences, and 717-, 559-, 353-, and 258-bp, respectively, using primer sets with added 5' flaps. Amplicons were eluted from agarose gels, quantified, and sequenced to confirm primer-target specificity. Primers with and without 5'flaps were pooled and used in multiplex PCR assays. In both single-target and multiplex reactions, sensitive and specific amplification was achieved with both the novel and modified primers. Sequencing confirmed primer-target amplification specificity, and phylogenetic analysis demonstrated that primer sets amplified different haplotypes of the biotypes. The addition of 5' flaps resulted in primer sensitivity and amplification differences, which were correlated with changes in the primer sets' combined ΔG values. Using these primer sets in single-target or multiplex PCR allows for quick discrimination and specific identification of *B. tabaci* biotypes and *T. vaporariorum*, and the addition of 5'A/T-rich overhang sequences increases primer sensitivity and amplification when primer ΔG is not increased.

Introduction

The whitefly (Hemiptera; Aleyrodidae) species *Bemisia tabaci* (Gennadius) and *Trialeurodes vaporariorum* (Westwood) are agricultural pests and vectors of plant viruses infecting food, fiber, and ornamental crops globally. The greenhouse whitefly, *T. vaporariorum*, is a pest on over 850 plant species and a natural vector of some viruses within the *Crinivirus* (Secoviridae) and *Torradovirus* (Closteroviridae) genera (Duffus, 1965; Wintermantle, 2004; Amari et al., 2008; CABI, 2015). The sweetpotato whitefly, *B. tabaci*, is a highly polyphagous pest and an efficient vector of the economically devastating begomoviruses (Geminiviridae), as

well as some Criniviruses, Torradoviruses, Carlaviruses (Betaflexiviridae), and Ipomoviruses (Potyviridae) (Costa, 1976; Jones, 2003; Navas-Castillo et al., 2011). With its high fecundity and virus transmission capabilities, *B. tabaci* often imposes limitations on crop production in many parts of the world (Brown, 1994). *B. tabaci* is described currently as a species complex composed of over twenty morphologically indistinguishable but biologically and genetically distinct biotypes or sibling species (Brown et al., 1995; Brown, 2010; De Barro et al., 2011). In this paper, these variant groups are referred to as biotypes, though their classification as biotypes or sibling species is debated. Compounding the issues raised by the complexity of this species are the genetic variants, or haplotypes, classified within the different biotypes (Frohlich et al., 1999). This wide genetic variation within the *B. tabaci* species complex challenges accurate identification for both research purposes (Boykin and DeBarro, 2014) and agricultural biosecurity efforts (Boykin et al., 2012).

The *B. tabaci* species complex is an ongoing threat to national agricultural biosecurity as evidenced by the introduction of two invasive biotypes into the United States in the past three decades. The B biotype, designated as the silverleaf whitefly, *B. argentifolii* (Bellows and Perring) after elevation to species status, was introduced in the 1980s and discovered as a result of sudden, unprecedented economic losses in greenhouse ornamentals (Bellows et al., 1994; Brown, 1995). Over time, the B biotype outcompeted and replaced the endemic A biotype, making for more challenging pest management primarily because the B biotype has higher pesticide resistance and is more efficient vector of begomoviruses. The highly pesticide-resistant Q biotype, introduced in the early 2000s, also has become widespread throughout the U.S. (Dennehy et al., 2005). Identification and interception of these high-consequence biotypes at U.S. ports of entry could significantly enhance our ability to minimize the introduction of novel geminiviruses, which frequently recombine, resulting in new variants (Padidam et al., 1999).

One limiting factor in PCR detection-based diagnostics is primer sensitivity. To achieve the highest possible sensitivity, primers ideally should be designed with optimal reaction thermodynamics; however, to achieve high specificity, thermodynamically optimal primers are not always an option. A previous study showed that adding 12-nucleotide long A/T-rich overhang sequences (5' flaps) to the 5' primer terminus increases primer sensitivity and PCR yield under some circumstances (Afonina et al., 2007). A subsequent comparison of optimally and sub-optimally designed primers demonstrated the degree to which 5' flaps can improve primer sensitivity and amplicon yield, but it also showed that such enhancement did not occur in every case or combination of flap-modified primers and their targets (Arif and Ochoa-Corona, 2013). Questions regarding the amplification effects of adding custom-designed, short, or dissimilar 5' flaps to primer sets remain.

Timely, accurate, and sensitive identification of agricultural pests and pathogens is fundamental to plant health inspection efforts at U.S. ports of entry. The aim of this study was to enhance capability to identify high-impact whitefly types by designing and validating target-specific endpoint PCR primers for accurate discrimination among *B. tabaci* (B, Q, and New World A-like biotypes) and between each of them and *T. vaporariorum*. Protocols for both single-target and multiplex discrimination were developed. Additionally, 5' flaps were added to each primer set to test for increased primer-target sensitivity. Phylogenetic analysis of the tested haplotypes was conducted to assess their relatedness as well as the accuracy of the biotype-specific primers.

Materials and Methods

Whitefly Sources and DNA Extraction

Global reference haplotypes, which include whiteflies collected over time from various locations worldwide and preserved in 95% ethanol, were obtained from Dr. Judith Brown's (University of Arizona) collection for primer specificity testing (Table III.1). To extract DNA, single whiteflies were removed from ethanol storage and allowed to dry for approximately 2 minutes. Whiteflies were transferred individually to 5 µL drops of PrepGEM® Insect (ZyGEM™, Hamilton, NZ) DNA lysis buffer on Parafilm M® and homogenized with a plastic pestle with the aid of a dissection microscope. The 5 µL homogenate was added to the remaining 35 µL of extraction buffer in a 0.5 mL microcentrifuge tube and mixed thoroughly. Tubes were placed in a PTC-100 thermocycler (MJ Research, Watertown, MA), incubated at 75˚C for 15 minutes, followed by 95˚C for 5 minutes, spun briefly, and placed in -20˚C storage until PCR testing.

Primer Design

B. tabaci and *T. vaporariorum* mitochondrial cytochrome oxidase c I (mtCOI) gene sequences were used for target-specific endpoint PCR primer design (Table III.2). Alignments were made using ClustalX2 (Larkin et al., 2007), and consensus sequences for *B. tabaci* Q, B, and A-like biotypes, and *T. vaporariorum* were created using GeneDoc (Nicholas et al., 1997). *B. tabaci* biotype- and *T. vaporariorum*-specific consensus sequence alignments were opened in GeneDoc to identify regions of divergence and single nucleotide polymorphisms for type-specific primer design. Discriminatory forward and reverse primers were identified for each of the four targets, and primer stability, forward/reverse compatibility, and melting temperature were evaluated using Primer3 (Table III.2; Untergrasser et al., 2012). Primer-target specificity was assessed using NCBI BLASTn. Target PCR amplicons for the *B. tabaci* Q, B, and A-like, and *T. vaporariorum* primer sets were 712-, 550-, 329-, and 252-bp, respectively. For simplification, these primer sets will be referred to as the Q, B, A, and T sets, respectively, throughout the text.

Table III.1. Inclusivity and exclusivity panel of *Bemisia tabaci* haplotypes, *Trialeurodes vaporariorum*, and other insect and microorganism species tested with species- and biotype-specific primer sets. Plus (+) indicates positive amplification, and minus (-) indicates no amplification. Asterisks indicate haplotypes and species also tested in multiplex reactions.

		Whitefly Discriminatory Primer Set			
		Q Biotype	B Biotype	A Biotype	*T. vaporariorum*
Whitefly Haplotypes and Panel Species	India39*	+	-	-	-
	BurkinaFaso2*	+	-	-	-
	Sudan2	+	-	-	-
	Mexico21*	+	-	-	-
	Israel35*	+	-	-	-
	Israel51	+	-	-	-
	Spain76	+	-	-	-
	Spain77*	+	-	-	-
	Spain5*	+	-	-	-
	Spain7*	+	-	-	-
	ControlQ*	+	-		-
	Argentina24*	+	-	-	-
	Argentina25	+	-	-	-
	Almeria16*	+	-	-	-
	USA55*	-	+	-	-
	USA58	-	+	-	-
	USA60*	-	+	-	-
	USA69	-	+	-	-
	USA74	-	+	-	-
	Brazil5*	-	+	-	-
	Brazil6	-	+	-	-
	Egypt23	-	+	-	-
	Mexico31*	-	+	-	-
	Israel34	-	+	-	-
	Israel53*	-	+	-	-
	Turkey10*	-	+	-	-
	Turkey19	-	+	-	-
	ColonyB*	-	+	-	-
	EN21*	-	+	-	-
	USA4	-	+	-	-
	USA7	-	+	-	-
	USA10*	-	+	-	-
	USA72	-	+	-	-
	USA79	-	+	-	-
	Egypt32	-	+	-	-
	SP21	-	+	-	-
	BIO6	-	+	-	-
	BIO17	-	+	-	-
	BIO31	-	+	-	-
	EN9*	-	+	-	-

Table III.1 (cont.). Inclusivity and exclusivity panel of *Bemisia tabaci* haplotypes, *Trialeurodes vaporariorum*, and other insect and microorganism species tested with species- and biotype-specific primer sets. Plus (+) indicates positive amplification, and minus (-) indicates no amplification. Asterisks indicate haplotypes and species also tested in multiplex reactions.

Whitefly Haplotypes and Panel Species	Q Biotype	B Biotype	A Biotype	*T. vaporariorum*
EN11	-	+	-	-
EN24	-	+	-	-
Mexico25*	-	-	+	-
ColonyA*	-	-	+	-
BIO13*	-	-	+	-
BIO19*	-	-	+	-
BIO24*	-	-	+	-
BIO26*	-	-	+	-
EN22*	-	-	+	-
IN69	-	-	-	-
Cameroon7*	-	-	-	-
EN2	-	-	-	-
Sudan1*	-	-	-	-
IN70	-	-	-	-
India33*	-	-	-	-
EN3*	-	-	-	-
Canada026*	-	-	-	+
Canada288*	-	-	-	+
Canada094*	-	-	-	+
Canada088*	-	-	-	+
MexicoTv*	-	-	-	+
*Trialeurodes abutilonea**	-	-	-	-
*Aleurotrachelus sp.**	-	-	-	-
Macrosteles quadrilineatus	-	-	-	-
Exitianus exitiosus	-	-	-	-
Endria inimica	-	-	-	-
Balclutha rubrostriata	-	-	-	-
Diuraphis noxia	-	-	-	-
Myzus persicae	-	-	-	-
Schizaphis graminum	-	-	-	-
Lepinotus reticulatus	-	-	-	-
Amblyomma americanum	-	-	-	-
Pristionchus pacificus	-	-	-	-
Fusarium proliferatum	-	-	-	-
Puccinia emaculata	-	-	-	-
Spiroplasma kunkelii	-	-	-	-
Salmonella enterica	-	-	-	-
Canna yellow streak virus	-	-	-	-

Table III.2. *Bemisia tabaci* biotype-discriminating and *Trialeurodes vaporariorum* species-specific primer sets with product and design specifications. Asterisks denote primers with 5' A/T-rich overhang sequences, and bold primer bases indicate these added 5' modifications.

Species and Biotype	Orientation	Primer Name	Primer Sequence 5'-3'	Product Size (bp)	TM (°C)	ΔG (kcal/mol)	Alignment Accession Numbers
Bemisia tabaci Q	Forward	QF	GTTTCTCATTTAATTAGCAGC	712	51.4	0.9	AF342773.1, DQ133377.1, AF342775.1, AY057138.1, DQ133378.1, AF342769.1, AF342776.1, DQ302946.1
	Reverse	QR	TGCTTACACCAAGCTTAAATCTTACTAAA		61.1	0.0	
	Forward	QF-f*	**TCATCATCA**TCATTTAATTAGCAGC	717	60.8	1.0	
	Reverse	QR	TGCTTACACCAAGCTTAAATCTTACTAAA		61.1	0.0	
B. tabaci B	Forward	BF	TATTTCACTTCAGCCACTATAA	550	52.2	0.3	DQ133375.1, AY057123.1, AF340215.1, AF340216.1, AM180064.1, AJ550176.1, DQ133372.1, DQ133373.1, AF321927.1, AY057140.1
	Reverse	BR	GCTTAAATCTTACTAACCGCAG		55.4	1.0	
	Forward	BF-f*	**TCATAA**TATTTCACTTCAGCCACTATAA	559	58.8	0.7	
	Reverse	BR-f*	**TCA**GCTTAAATCTTACTAACCGCAG		60.3	1.0	
B. tabaci A-like	Forward	AF	TAAGTTTAGACCCCTAGTTCTC	329	53.9	0.9	AY057122.1, AY057124.1, DQ133368.1, AY057125.1, AF342770.1, DQ133376.1, DQ133371.1, AY057128.1, AY057126.1, AF342772.1, AY057133.1, AY057127.1
	Reverse	AR	CAGAATACCGACGAGGT		51.4	0.0	
	Forward	AF-f*	**AATAAATCATAA**TAAGTTTAGACCCCTAGTTCTC	353	61.2	0.9	
	Reverse	AR-f*	**AATAAATCATAA**CAGAATACCGACGAGGT		63.2	0.0	
Trialeurodes vaporariorum	Forward	TvF	TGTCATTTAATCCCCTTACTTC	252	55.0	1.0	AF110708.2, JX841216.1, JX841220.1, JQ995231.1, LN614547.1, KJ475452.1, HE863766.1, JF693934.1, KF991608.1, KC843064.1
	Reverse	TvR	ACAAAACTGGGAAAGAAGAAG		56.7	0.0	
	Forward	TvF-f*	**TCA**TGTCATTTAATCCCCTTACTTC	258	59.7	1.0	
	Reverse	TvR-f*	**TCA**ACAAAACTGGGAAAGAAGAAG		61.0	0.0	
B. tabaci General	Forward	Bt Cox 1-628F	GATCGAAATTTTAATAGATCTTTTTATGATCC	1000	63.5	1.0	
	Reverse	Bt Cox 2-1629R	TGTTCTATTGTAAAACTAGCACTATTTTG		62.1	0.9	

To test for increased primer sensitivity and target amplification, 5' flaps were incorporated at the 5' termini of each primer (Table III.2; Afonina et al., 2007). Different 5' flap types were added to each of the four novel primer sets. The full 12-nucleotide A/T-rich flap (AATAAATCATAA) designed by Afonina et al. (2007) was incorporated at the 5' termini of the A primers. Questions posed by Arif and Ochoa-Corona (2013) regarding the effects of adding custom-designed, short, or dissimilar 5' flaps were addressed by adding different 5' flap types to the Q, B, and T primer sets (Table III.2). The Q primer set was synthesized with a customized 5' flap (5'-TCATCATCA-primer sequence-3') on the forward primer only, and no 5' flap was added to the reverse primer. For the B biotype set, dissimilar customized 5' flaps were added to the forward (TCATAA) and reverse (TCA) primers. Customized, short 5' flaps (TCA) were added to the *T. vaporariorum* set. Thus, each primer set had a different type of 5' flap added, which could be compared to the originally designed set to determine the amplification effect of that particular 5' flap type when added to a given primer set. PCR amplicons derived from primer sets incorporating 5' flaps were 717-, 559-, 353-, and 258-bp for the Q, B, and A-like biotypes, and *T. vaporariorum*, respectively. These primer sets will be referred to as the Q-flap, B-flap, A-flap, and T-flap sets, respectively. All primers were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). The Gibbs free energy (ΔG) of each primer (Table III.2) was assessed using mFold software, which predicts DNA and RNA folding structures for input sequences and their ΔG at defined annealing temperatures and ionic conditions (Zuker, 2003). See Appendix 1 for mFold results for each primer.

Single-target PCR Validation

An optimal annealing temperature was determined for each of the eight primer sets by gradient PCRs performed using 10 µL of 2X GoTaq Green Master Mix (Promega, Madison, WI), 6 µL nanopure water, 1 µL of 5 µM forward primer, 1 µL of 5 µM reverse primer, and 2 µL of 20

ng/µL target DNA for a 20 µL reaction. Thermocycling conditions were 95˚C for 3 minutes; 35 cycles of 94˚C for 15 seconds, 53-65˚C for 30 seconds, and 72˚C for 45 seconds; and 72˚C for 3 minutes. Reactions were performed in a Biometra TProfessional Thermocycler (Göttingen, Germany). Gradient PCR products were separated by electrophoresis in a 1.5% agarose gel with a 1 Kb Plus DNA Ladder (Invitrogen, Carlsbad, CA).

The specificities of the eight primer sets were tested at the optimal annealing temperature of 57ºC. Positive control reactions consisted of the whitefly type-specific primer set and positive control target DNA, while negative control reactions contained the target type-specific primer set and nanopure water. Each validation PCR, setup and run as described above, was replicated three times using different samples of target whitefly DNA at 40 ng per reaction. PCR products were separated by electrophoresis in a 2% agarose gel with 4 µL SYBR® Safe DNA Gel Stain (Invitrogen) per 100 mL gel and 4 µL of Low DNA Mass Ladder (Invitrogen) in 1X TAE. Resulting amplicons were eluted from agarose gels using Quantum Prep Freeze 'N Squeeze Spin Columns (Bio-Rad, Hercules, CA), and the products were quantified three times using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE). Purified DNA was submitted to the Oklahoma State University Biochemistry and Molecular Biology Recombinant DNA and Protein Core Facility for sequencing with an ABI 3730 DNA Analyzer (Applied Biosystems^{TM}, Carlsbad, CA). Resulting sequences were trimmed and aligned using ClustalX2, and target specificity was confirmed by alignment using the NCBI BLASTn algorithm.

To further validate the specificity of designed primer sets for global reference haplotypes, PCR tests were carried out on DNA extractions (previously described) from *B. tabaci* collected worldwide (summarized in Table III.1). The *B. tabaci* Q, B, and A-like primer sets lacking 5' flaps were tested on over 120 *B. tabaci* individuals from 56 populations. Each primer set was

tested additionally on DNA extracted from a panel of related whitefly species as well as unrelated organisms to exclude the possibility of amplification from non-target species or contamination (Table III.1). PCR mixes were prepared with crude DNA lysates and reaction conditions were performed as described above. Products were separated by electrophoresis in a 1.5% agarose gel.

Multiplex PCR Amplification

Primer sets with and without 5' flaps were pooled separately (Q, B, A, and T pool versus Q-flap, B-flap, A-flap, and T-flap pool) at equal concentrations of each primer to test primer set compatibility and specificity in multiplex PCR reactions. Multiplex reactions contained 10 µL of 2X GoTaq Green Master Mix, 0.125 µM concentrations of each primer, 1 µL of 25 mM $MgCl_2$, 1 µL of 20 ng/µL or crude extract target DNA, and an appropriate volume of nanopure water to bring the reaction volume to 20 µL. Reactions were run in a Biometra TProfessional Thermocycler, and conditions used were 95˚C for 5 minutes; 35 cycles of 94˚C for 30 seconds, 58˚C for 1 minute, and 72˚C for 1 minute; and 72˚C for 7 minutes. In addition to the above primer set pooling and testing, the Q and B primers were pooled and tested with the A-flap and T-flap primers. These tests contained the same reagent concentrations and reaction conditions. Multiplex PCR products were separated by electrophoresis in a 2% agarose gel with a 1 Kb Plus DNA Ladder.

Phylogenetic Analysis

To confirm the identity of each haplotype tested in single-target and multiplex reactions and to evaluate their phylogeny, *B. tabaci* general primers previously designed in the Brown lab (University of Arizona) amplified a diagnostic region of the mtCOI – mtCOII genes (Table III.2). The products were cleaned and sequenced directly at the University of Arizona Genetic Core laboratory using an ABI 3730 DNA Analyzer. The resulting sequences were aligned, cleaned, and

trimmed using MEGA5 (Tamura et al., 2011), and one sequence from each tested haplotype was used to generate a Maximum Likelihood phylogenetic tree. Sequences were aligned using MUSCLE (Edgar, 2004). Phylogenetic analyses were conducted using MEGA7 (Kumar et al., 1993), and relationships were inferred based on the Tamura-Nei model (Tamura and Nei, 1993). The tree was drawn to scale with branch lengths measured in the number of substitutions per site. *T. vaporariorum* was used as an outgroup.

Results

Single-target PCR Amplification

Gradient PCRs demonstrated that 57ºC was an optimal annealing temperature for evaluating each primer set (Figure III.1). In the initial triplicate validation reactions, 40 ng of target and non-target DNA was tested. Each primer set designed herein specifically amplified its corresponding target biotype or species in single target reactions and did not amplify any non-target biotypes or other species (Figure III.2; Table III.1). Amplicons were produced for the target whitefly primer sets and corresponding 5' flap modified primer sets tested with whitefly target DNA as well as for the positive control reactions. No amplification was observed in any of the non-target or negative control reactions.

Use of the 5' flap modified primer sets for PCR did not consistently improve product intensity, compared to the use of the designed sets. Bands produced by the Q-flap and B-flap sets were less intense in each replicate than those produced by the no-flap sets (Figure III.2). Adding the customized 5' flaps to these primer sets decreased the target-sensitivity at the cycling conditions used. However, use of the A-flap and T-flap primers resulted in greater band intensities than were produced using their no-flap counterparts (Figure III.2). In reactions with

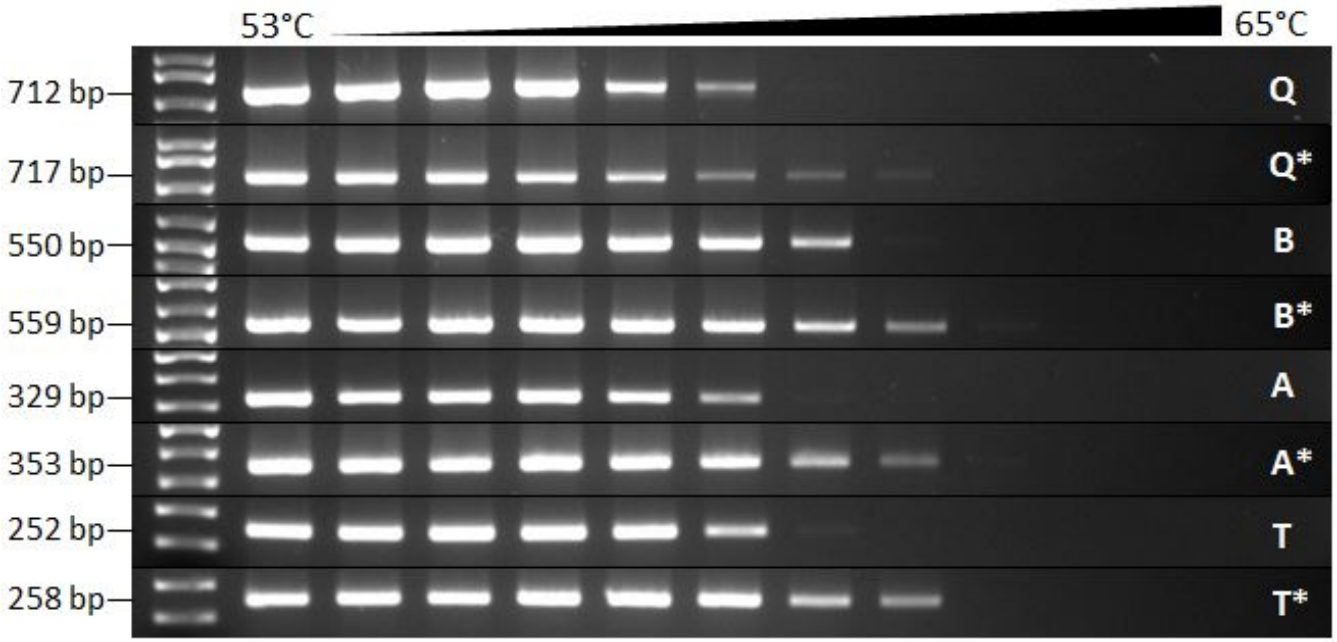


Figure III.1. Gradient PCR amplification from 53ºC to 65 ºC using *Bemisia tabaci* biotype-specific (Q, B, and A) and *Trialeurodes vaporariorum* species-specific (T) primer sets. Asterisks denote primer sets with 5' A/T-rich overhang sequences. 57°C was chosen as an optimal annealing temperature for each primer set. The fifth lane after the ladder represented assays performed at an annealing temperature of 57.33°C.

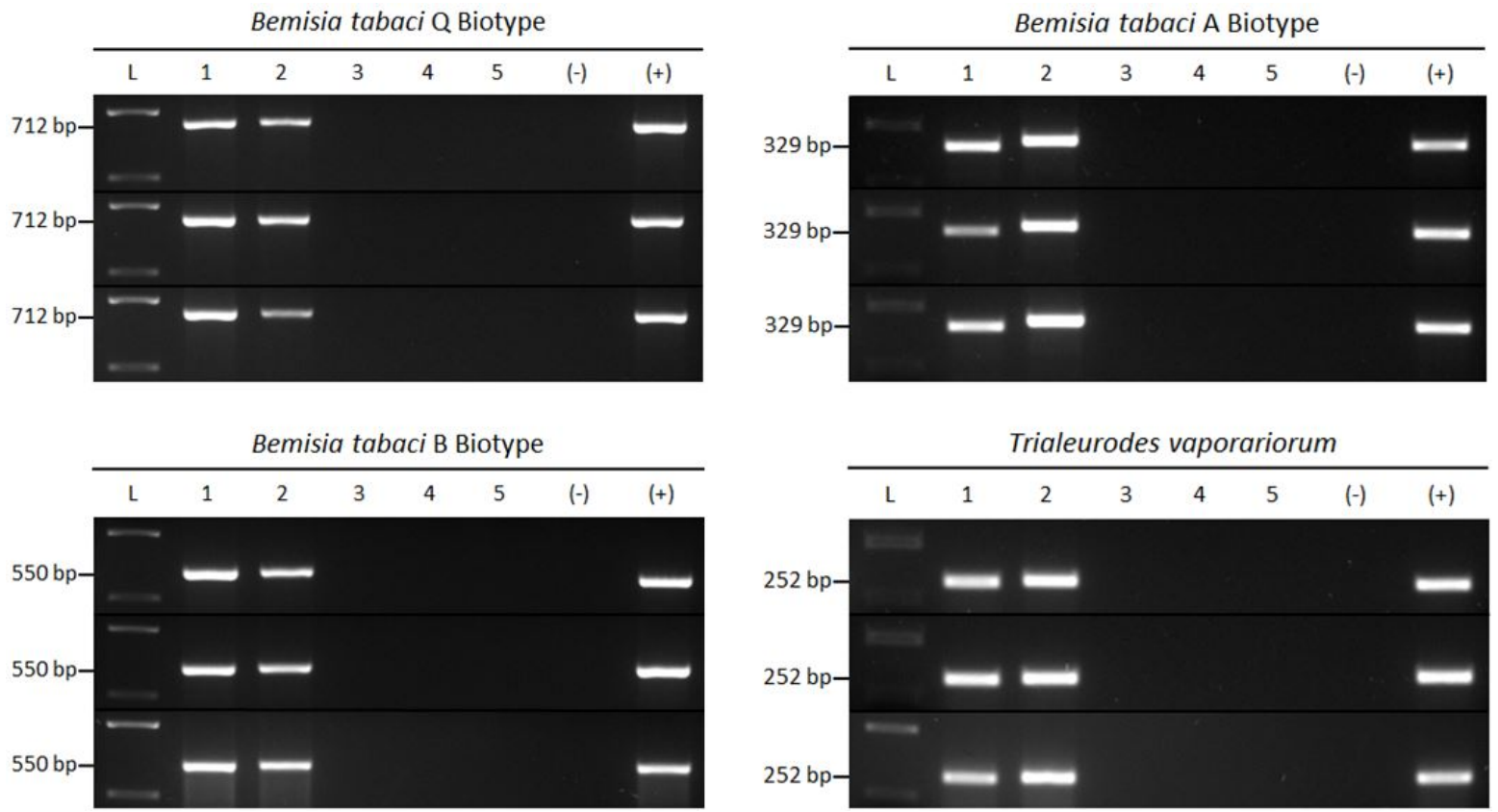


Figure III.2. Replicate validation PCRs using *Bemisia tabaci* biotype- and *Trialeurodes vaporariorum* species-specific primer sets. Endpoint PCR amplification of targeted discriminatory segments of *B. tabaci* and *T. vaporariorum* mtCOI DNA in 2% agarose gels electrophoresed with a Low DNA Mass Ladder in 1X TAE. Lane 1 contains products from designed primer sets and target DNA. Lane 2 contains corresponding primer sets with 5' A/T-rich overhang sequence modifications and target DNA. Lanes 3-5 contain PCR assays with designed primer sets on non-target DNA; for Q gels: B, A, and T DNA; for B: Q, A, and T; for A: Q, B, and T; and for T: Q, B, and A. Negative and positive controls follow.

these two targets, target-sensitivity and amplification were increased by the addition of 5' flaps. NanoDrop quantification of resulting amplicons verified these results (Figure III.3), which were significant (unpaired *t* test). The Q-flap and B-flap primer sets significantly decreased whitefly target amplification compared to the Q and B sets by an average of 44.9 ($p < 0.019$) and 25.2 ($p < 0.028$) percent, respectively (Figure III.3). Whereas, the A-flap and T-flap sets significantly increased amplification compared to the A and T sets by an average of 54.5 ($p < 0.011$) and 36.6 ($p < 0.027$) percent, respectively (Figure III.3).

Sequencing of eluted products from the triplicate validation reactions showed that only the target biotype or species was amplified by each primer set; non-targets were not amplified. The resulting forward and reverse sequences were compared and consensus sequences were determined. Accounting for uncertain basecalling quality at read ends, consensus sequences were trimmed to 614-, 540-, 298-, and 249-bp in length for the Q, B, and A-like biotypes, and *T. vaporariorum*, respectively. Each sequence aligned with sequences of their target type with 100% query coverage and 100% identity against the NCBI GenBank nucleotide database.

Further validation of the primers on haplotypes collected globally demonstrated that the primers amplify different haplotypes classified within the target biotypes (Table III.1). Individuals from 14 different populations were amplified by the Q primers, 29 different populations by the B primers, and 7 different populations by the A primers with no amplification by non-target biotype sets. Additionally, 6 different populations of *B. tabaci* were not amplified by any primer set. Sequencing of tested individuals confirmed that each haplotype was amplified by its corresponding biotype-specific primers. Those haplotypes that were not amplified by any primer set were confirmed as non-Q, -B, or -A-like biotypes of *B. tabaci*.

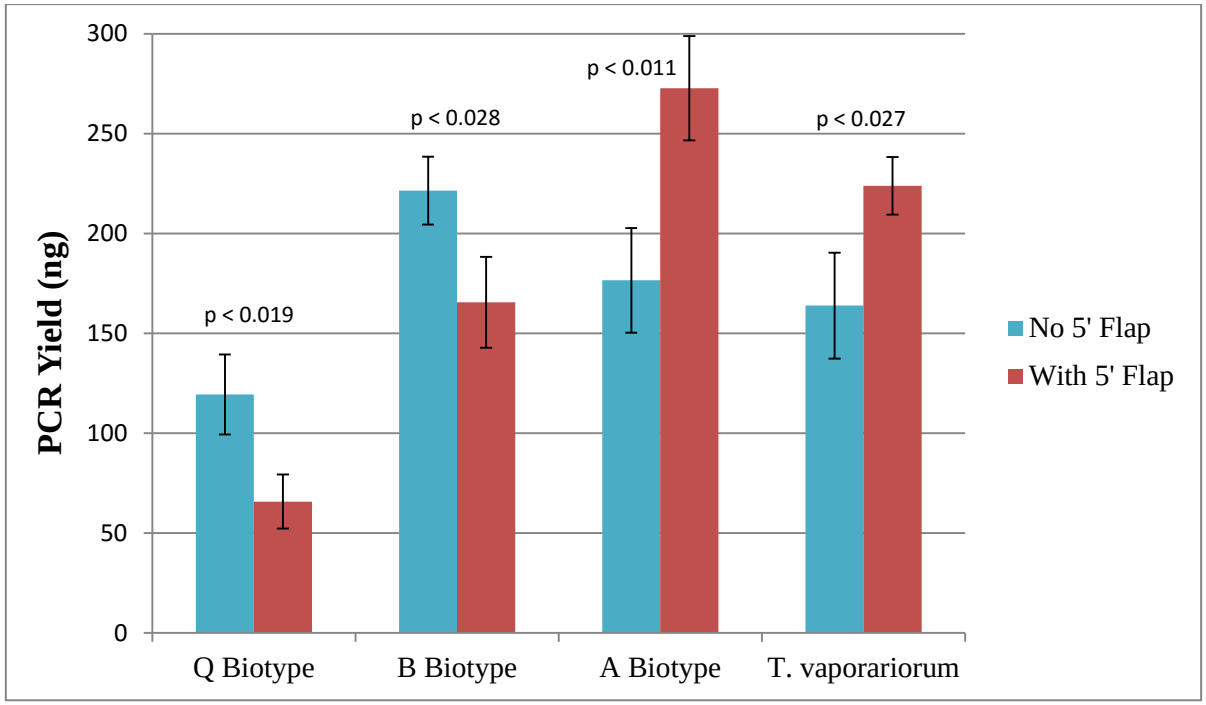


Figure III.3. Average yields and standard deviations of three PCR replicates conducted with *Bemisia tabaci* biotype- and *Trialeurodes vaporariorum* species-specific primer sets with and without 5' A/T-rich overhang sequences (5'flap). Significant differences were observed between average yields of primer sets with versus without 5' flaps (unpaired *t* test; Q biotype: $p < 0.019$; B biotype: $p < 0.028$; A biotype: $p < 0.011$; *T. vaporariorum*: $p < 0.027$).

Multiplex PCR Amplification

Primers designed for discriminating species and biotypes in single-target reactions specifically amplified their target species or biotype when combined in multiplex reactions. Initial multiplex assays using separately pooled primers, with and without 5' flaps (Q, B, A, and T pool versus Q-flap, B-flap, A-flap, and T-flap pool), resulted in non-specific amplification of the Q DNA only and in multi-target reactions (results not shown). For this reason and because the Q and B primer sets were more sensitive than the Q-flap and B-flap primers in single-target reactions, the Q and B primer sets were mixed and tested with the A-flap and T-flap sets. When this primer pool (Q and B with A-flap and T-flap) was used in multiplex reactions, target-specific amplification without non-specific amplification was achieved (Figure III.4). Different haplotypes were tested in these reactions (Figure III.4; Table III.1) and demonstrated the biotype target-sensitivity and repeatability of the multiplex amplification protocol.

Phylogenetic Analysis

The *B. tabaci* haplotypes tested were clustered into five distinct clades representing the Q, B, and A-like biotypes and two non-target biotype groups (Figure III.5). Each haplotype that was amplified using biotype-specific primers grouped into the Q, B, or A-like group clade. The B and Q groups clustered as sister clades. Non-target biotypes clustered into two clades, one as a sister clade to the A-like group, and the other as an outgroup to the target biotypes and the other non-target clade. *T. vaporariorum* was the outgroup taxon to all *B. tabaci* haplotypes.

Discussion

A rapid PCR protocol for discrimination of morphologically indistinguishable *B. tabaci* biotypes and *T. vaporariorum,* using specifically designed endpoint PCR primers, was designed

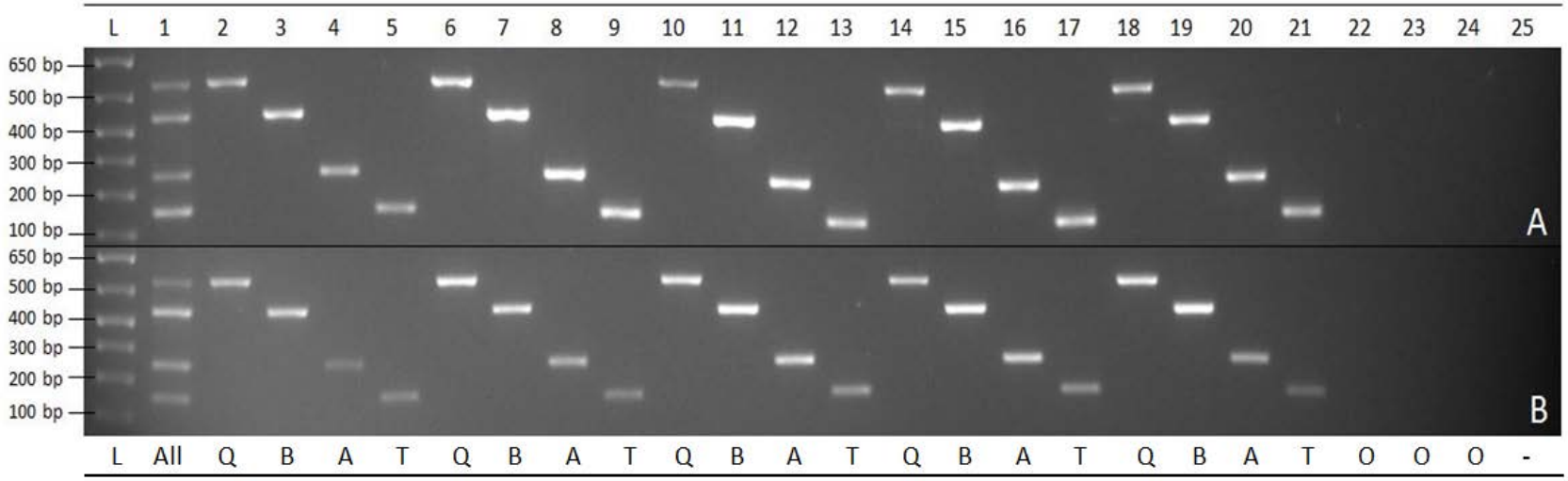


Figure III.4. Multiplex PCR amplification using mix of *Bemisia tabaci* biotype- and *Trialeurodes vaporariorum* species-specific primer sets with and without 5' A/T-rich overhang sequences on *B. tabaci* haplotype and *T. vaporariorum* DNA in 2% agarose gels electrophoresed with a 1 Kb Plus Ladder in 1X TAE. Letters indicated below the gel image correspond lane contents: (L) ladder; (All) all four targets tested; (Q) Q biotype tested; (B) B tiotype tested; (A) A biotype tested; (T) *T. vaporariorum* tested; (O) other biotype tested; (-) negative control. Lanes 1A and 1B contain amplicons from all four targets with Q biotype at 712 bp, B biotype at 550 bp, A-like biotype group at 353 bp, and *T. vaporariorum* at 258 bp. Lanes 2A – 24A contain the following *B. tabaci* haplotype or *T. vaporariorum* individuals: ControlQ, ColonyB, ColonyA, Canada026, India39, USA55, Mexico25, Canada288, BurkinaFaso2, Brazil5, EN22, Canada094, Mexico21, Israel53, BIO24, Canada088, Israel35, Turkey10, BIO13, MexicoTv, Cameroon7, Sudan1, and *T. abutilonea*. Lanes 2B – 24B contain the following individuals: Spain5, EN21, BIO19, Canada026, Spain7, USA10, BIO26, Canada288, Argentina24, Mexico31, Mexico25, Canada094, Almeria16, USA60, EN22, Canada088, Spain77, EN9, BIO24, MexicoTv, EN3, India33, and *Aleurotrachelus sp.* Lanes 25A and 25B contain negative (NTC; non template target) controls.

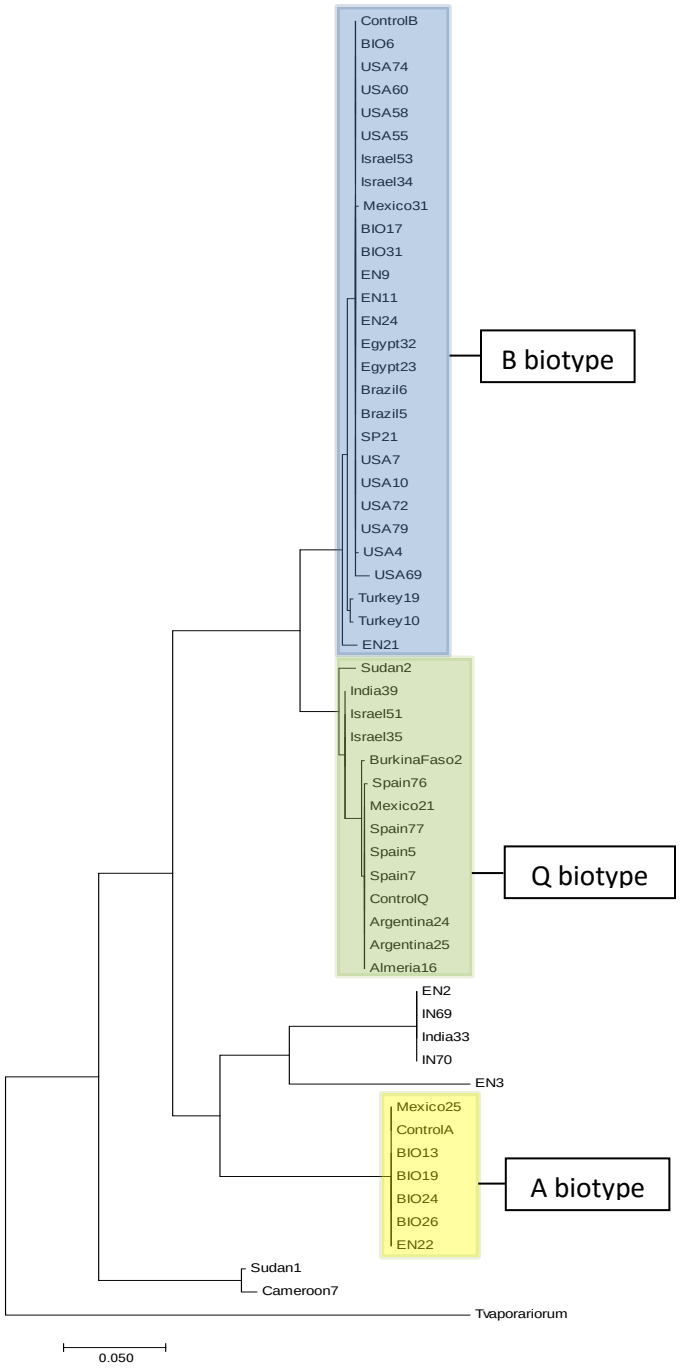


Figure III.5. Phylogenetic relationships of tested *Bemisia tabaci* populations represented by Maximum Likelihood analysis using the Tamura-Nei model with *Trialeurodes vaporariorum* as an outgroup. Sequences of the 1000-bp mtCOI-mtCOII genes region were amplified with *B. tabaci* Bt Cox 1-628F and Bt Cox 2-1629R primers. Populations highlighted in the blue box correspond to individuals of the B biotype amplified only by novel B biotype primers; Populations highlighted in the green box correspond to individuals of the Q biotype amplified only by novel Q biotype primers; Populations highlighted in the yellow box correspond to individuals of the A biotype amplified only by the novel A biotype primers.

and validated in this study. The specificities of each primer set to their targets were tested and validated on over 120 individuals of various haplotypes of the B, Q, and A-like groups, as well as non-target biotypes, in single-target reactions. The primers are specific for their targets at the optimal 57°C annealing temperature, allowing for amplification of their targets at a minimal 1°C range around this temperature to account for thermocycler differences. Phylogenetic analysis of the haplotypes assayed supported the high degree of specificity of these primers. Sequences from all haplotypes resulting in bands from a particular primer set clustered within their particular biotype clades, and no bands were observed from non-target haplotypes within non-Q, B, or A-like group phylogenetic clades.

Use of primers modified with the insertion of customized 5' flaps resulted in either increased or decreased sensitivity in single-target reactions, depending on the primer set. In some cases reported previously (Afonina et al., 2007; Arif and Ochoa-Corona, 2013), 5' flaps were shown to increase primer sensitivity, resulting in higher yield and brighter, more intense bands in endpoint PCR and increased fluorescence in qPCR. However, in other cases, 5' flap addition has not led to consistent improvement, and Arif and Ochoa-Corona (2013) concluded that primer sensitivity enhancement by flap addition is dependent on the primer sequence and is likely to be effective primarily in cases in which existing primers are characterized by sub-optimal thermodynamics.

Our results showed that both a full 12-nucleotide 5' flap described by Afonina et al. (2007) and short, customized matching 5' flaps, added to both forward and reverse primers, significantly increased primer sensitivity, while customized dissimilar forward and reverse 5' flaps and adding a customized 5' flap to only one primer in the set did not increase primer sensitivity. These data appear to correlate with the combined ΔG of each primer set. A ΔG value equal to zero represents a primer at equilibrium at the specified annealing temperature and ionic conditions. For both the Q and B sets, the ΔG of one primer in the set was increased by adding 5'

flaps, while in the A and T sets the ΔGs remained the same. This slight increase (0.05 and 0.2 kcal/mol increase for Q to Q-flap and B to B-flap, respectively) likely reduced the reaction thermodynamic stability resulting in less target amplification. Primers should be evaluated with mFold prior to the addition of 5' flaps. Considering these results, in the case that the ΔG of either primer or both primers in the set is increased, potentially, both target sensitivity and amplification will decrease. Conversely, if the ΔG remains the same or perhaps lessens, sensitivity and amplification may increase. Primer sets made more or less stable by 5' flap addition, their effects on primer sensitivity and amplification, and the effects of primer ΔG on target amplification should be tested further.

When using these primers in multiplex, separate pooling of primers with and without 5' flaps resulted in some non-specific amplification in both reaction types. However, multiplexed PCR reactions using a mix of primers with and without 5'flaps (Q and B primer sets with the A-flap and T-flap sets) resulted in target-specific amplification with no non-specific amplification. The four primer sets (Q, B, A-flap, and T-flap) used in these multiplex assays were shown also to be more sensitive than their counterparts (Q-flap, B-flap, A, and T) in single-target reactions. Specific multiplex amplification of a range of *B. tabaci* haplotypes and *T. vaporariorum* was consistent and repeatable in reactions incorporating this specific primer mix.

In a previous study, Q, B, and A biotype-specific primer sets were designed and used in combined primer set reactions, demonstrating multiplex amplification of these three *B. tabaci* types (Shatters et al., 2009). This study provides new single-target and multiplex primer sets and discrimination protocols for Q, B, the A-like group *B. tabaci* biotypes, along with *T. vaporariorum*. It also explores the use of primer modifications that have been reported to increase PCR sensitivity and amplification. Use of the multiplex protocol described herein provides rapid identification of the two highly problematic whitefly species, *B. tabaci* and *T. vaporariorum*, and

specific discrimination among two high-consequence *B. tabaci* biotypes, B and Q, as well as the native A and A-like biotypes.

In the absence of consistent and convenient biotype-discrimination tools at ports of entry, future introductions of other exotic haplotypes, viruses, or novel vector-virus combinations are likely. Incorporation of discriminatory assay techniques, such as those described here, at ports of entry could facilitate interventions and minimize further introductions of exotic haplotypes and any plant viruses they might harbor.

REFERENCES

Afonina, I., Ankoudinova, I., Mills, A., Lokhov, S., Huynh, P., and Mahoney, W. 2007. Primers with 5′ flaps improve real-time PCR. BioTechniques 43:770–772.

Arif, M., and Ochoa-Corona, F. M. 2013. Comparative assessment of 5′ A/T-rich overhang sequences with optimal and sub-optimal primers to increase PCR yields and sensitivity. Molecular biotechnology 55:17-26.

Amari, K., Gonzalez-Ibeas, D., Gómez, P., Sempere, R. N., Sanchez-Pina, M. A., Aranda, M. A., Diaz-Pendon, J. A., Navas-Castillo, J., Moriones, E., Blanca, J. and Hernandez-Gallardo, M. D. 2008. *Tomato torrado virus* is transmitted by *Bemisia tabaci* and infects pepper and eggplant in addition to tomato. Plant Disease, 92:1139-1139.

Bellows, T. S., Perring, T. M., Gill, R. J., Headrick, D. H. 1994. Description of a species of *Bemisia* (Homoptera: Aleyrodidae). Annals of the Entomological Society of America 87:195-206.

Boykin, L. M., Armstrong, K. F. Kubatko, L. and De Barro, P. 2012. Species delimitation and global biosecurity. Evolutionary bioinformatics online 8:1.

Boykin, L. M., and De Barro, P. J. 2014. A practical guide to identifying members of the *Bemisia tabaci* species complex: and other morphologically identical species. Frontiers in Ecology and Evolution 2:45.

Brown, J. K. 1994. Current status of *Bemisia tabaci* as a plant pest and virus vector in agro-ecosystems worldwide. FAO Plant Prot. Bull. 42:3-32.

Brown, J. K. 2010. Phylogenetic biology of the *Bemisia tabaci* sibling species group. *Bemisia*: Bionomics and Management of a Global Pest. Springer Netherlands 31-67.

Brown, J. K., Frohlich, D. R. and Rosell, R. C.. 1995. The sweetpotato or silverleaf whiteflies: biotypes of *Bemisia tabaci* or a species complex? Annual review of entomology 40:511-534.

CABI. 2015. *Trialeurodes vaporariorum* (whitefly, greenhouse). In: Invasive Species Compendium. Wallingford, UK: CAB International. http://www.cabi.org/isc/datasheet/54660

Costa, A. S. 1976. Whitefly-transmitted plant diseases. Annual Review of Phytopathology 14:429–49

De Barro, P. J., Liu, S. S., Boykin, L. M. *and* Dinsdale, A. B. 2011. *Bemisia tabaci*: a statement of species status. Annual Review of Entomology 56*:*1–19.

Dennehy, T. J., B. A. Degain, V. S. Harpold, J. K. Brown, S. Morin, J. A. Fabrick, F. J. Byrne, and R. L. Nichols. 2005. New challenges to management of whitefly resistance to insecticides in Arizona. The University of Arixona Co-operative Extension Report, Tucson, AZ.

Duffus, J. E. 1965. *Beet pseudo-yellows virus*, transmitted by the greenhouse whitefly, *Trialeurodes vaporariorum*. Phytopathology 55:450-453.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-97.

Frohlich, D. R., I. Torres-Jerez, I. D. Bedford, P. G. Markham, and J. K. Brown. 1999. A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. Molecular ecology 8:1683-1691.

Jones D. R. 2003. Plant viruses transmitted by whiteflies. Eur. J. Plant Pathol. 109:195–219.

Kumar, S., Stecher, G., and Tamura, K. 2015. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution, submitted.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948

Navas-Castillo, J., Fiallo-Olivé, E., and Sánchez-Campos, S. 2011. Emerging virus diseases transmitted by whiteflies. Annual review of phytopathology 49:219-248.

Nicholas, K. B., Nicholas H. B. Jr., and Deerfield, D. W. II. 1997. GeneDoc: Analysis and Visualization of Genetic Variation, EMBNEW.NEWS 4:14.

Padidam, M., Sawyer, S., and Fauquet, C. M. 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265:218-225.

Shatters, R. G., Powell, C. A., Boykin, L. M., Liansheng, H., and McKenzie, C. L. 2009. Improved DNA barcoding method for *Bemisia tabaci* and related Aleyrodidae: development of universal and *Bemisia tabaci* biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. Journal of economic entomology 102: 750-758.

Tamura, K. and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10:512-526.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28:2731-2739.

Untergrasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., Rozen, S. G. 2012. Primer3 - new capabilities and interfaces. Nucleic Acids Research 40:e115

Wintermantel, W. M. 2004. Emergence of greenhouse whitefly (*Trialeurodes vaporariorum*) transmitted criniviruses as threats to vegetable and fruit production in North America. APSnet Feature Story June 2004.

Zuker, M. (2003). mFold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31, 3406–3415.

CHAPTER IV

MELTING TEMPERATURE ANALYSIS AND HELICASE DEPENDENT AMPLIFICATION DISCRIMINATION OF *BEMISIA TABACI* BIOTYPES AND *TRIALEURODES VAPORARIORUM*

Abstract

The tobacco or sweetpotato whitefly, *Bemisia tabaci* (Gennadius) (Hemiptera; Aleyrodidae), and greenhouse whitefly, *Trialeurodes vaporariorum* (Westwood), are highly problematic plant pests with worldwide distributions. Identification of whitefly species is typically accomplished by observation of distinct morphological characters; however, the discrimination of *B. tabaci* biotypes is dependent on molecular techniques based on genetic differences. New assays were designed for the detection of *B. tabaci* Q, B, and A biotypes, and *T. vaporariorum.* Specific primer sets were designed for amplification of the four targets in end-point PCR, melting temperature analysis, and helicase dependent amplification assays. Primer specificities were validated using end-point PCR then tested in melting temperature analysis and helicase dependent amplification. *B. tabaci* Q, B, and A biotypes, and *T. vaporariorum*-targeted primer sets successfully amplified different population samples of their target whitefly type, providing three new discrimination assays for whitefly identification.

Introduction

World food, fiber, and greenhouse crops are damaged annually by the ubiquitous whitefly pest *Bemisia tabaci* (Gennadius) (Hemiptera; Aleyrodidae), vector of begomoviruses. *B. tabaci* was first reported as a pest of tobacco in Greece in 1889 (Gennadius, 1889). Over the next century, the species was described in other parts of the world (United States, Brazil, Taiwan, etc.) as a pest of other crop species and given various taxonomic names (Quaintance, 1900; Mound and Halsey, 1978). In the 1980s, *B. tabaci* gained global attention after being exported to many countries worldwide on poinsettia from the Middle East and subsequently becoming a limiting factor in crop production in many areas where it was previously manageable (Brown, 1994). The establishment of the more invasive *B. tabaci* prompted an examination into the composition of this species that continues today.

B. tabaci is currently considered a species complex composed of many (possibly over 35) cryptic species that have considerable genetic and biological differentiation (De Barro *et al.,* 2011; Liu *et al.,* 2012; Esterhuizen *et al.,* 2013; Barbosa *et al.,* 2014). Despite being morphologically indistinguishable, *B. tabaci* species variants (also referred to as biotypes) differ in biological characteristics including host range, fecundity, pesticide resistance, behavior, ability to induce plant physiological changes, and begomovirus transmission efficiency. These differences translate to *B. tabaci* biotypes posing varying degrees of agricultural risk. Two exotic, high-consequence biotypes, the B and Q biotypes from the Middle East and Mediterranean regions, respectively, were introduced and became established in the U.S. in the last three decades. The ability to correctly identify biotypes is of critical interest to entomologists working at U.S. ports of entry as well as to researchers and growers.

Because of the inability to distinguish species variants morphologically, which is common protocol for pest identification at U.S. ports of entry, biotype differentiation has relied

on molecular techniques. Differences in esterase profiles (Costa and Brown, 1991), allozymes (Perring et al., 1992), and RAPD-PCR (Gawell and Bartlett, 1993) demonstrated genetic variation among the first biotypes described in the U.S. (endemic A and exotic B). Trace-back of the B biotype to the Middle East was accomplished in a phylogeographic study of mitochondrial COI and 16S genes in 10 worldwide populations (Frohlich et al., 1999). Further studies based on genetic techniques, including AFLP (Cervera et al., 2000), RFLP (Ma et al., 2009), and microsatellites (De Barro, 2005) have further established genomic differences among biotypes and the use of such variation for biotype discrimination (Shatters et al., 2009; De Barro et al., 2011).

PCR protocols have been developed for the detection of specific biotypes. Multiplex PCR differentiation of the high-consequence B and Q biotypes with A (Shatters et al., 2009) and additionally with another important whitefly vector species, *Trialeurodes vaporariorum* (Westwood), (Andreason et al., in preparation) have provided quick discrimination methods with end-point PCR. A TaqMan real-time PCR assay has been developed for rapid B and Q biotype differentiation (Papayiannis et al., 2009), and loop-mediated isothermal amplification (LAMP) assays have also been developed for field-based B and Q identification (Adachi et al., 2010; Hsieh et al., 2012; Dickey et al., 2013). These real-time and isothermal detection strategies are, however, limited to detection of the B and Q biotypes only.

As the risk of inadvertent introduction of invasive biotypes, begomoviruses, or new virus-vector combinations remains, and the ability to detect and verify biotypes under experimentation or collected in the field is a persistent need, rapid discrimination strategies are necessary. The objectives of this study were to design primer sets that discriminately amplify the *B. tabaci* Q, B, and A biotypes and *T. vaporariorum* in melting temperature analysis (MTA) and helicase dependent amplification (HDA) assays. These assays were chosen because MTA presents a real-

time PCR-based amplification assay that negates the need for both post-assay gel electrophoresis and traditional qPCR amplicon quantification requirements for detection, and HDA provides an isothermal amplification option that does not require extensive primer design and optimization per target (as needed for LAMP) and is field-deployable (thermocylcing unnecessary). To my knowledge, neither method has been reported for identification of *B. tabaci* biotypes or *T. vaporariorum*.

Materials and Methods

Whiteflies and DNA Extraction

B. tabaci whiteflies collected over time from various locations worldwide and preserved in 95% ethanol were provided by Dr. Judith Brown (University of Arizona). *T. vaporariorum* were generously provided by Dr. John Dooley (USDA APHIS). DNA was extracted from single whiteflies of five different populations for each target type (Table IV.1) by removing individuals from ethanol storage, allowing them to dry for approximately 2 minutes, and transferring to 5 µL drops of PrepGEM Insect™ (ZyGEM, Hamilton, NZ) DNA lysis buffer on Parafilm$_{TM}$. Whiteflies were homogenized with a plastic pestle and the aid of a Wild M5 dissection microscope. The 5 µL homogenate was added to the remaining 35 µL of extraction buffer in a 0.5 mL microcentrifuge tube and mixed thoroughly. Tubes were placed in a PTC-100 thermocycler (MJ Research, Watertown, MA), incubated at 75˚C for 15 minutes, followed by 95˚C for 5 minutes, spun briefly, and placed in -20˚C storage until PCR testing.

Primer Design

B. tabaci and *T. vaporariorum* mitochondrial cytochrome oxidase c I (mtCOI) gene sequences were used for target-specific PCR primer design (Table IV.2). Alignments were made

Table IV.1. *B. tabaci* and *T. vaporariorum* populations assayed by end-point PCR (E), melting temperature analysis (M), and/or isothermal helicase dependent amplification (I) using primer sets designed for *B. tabaci* Q, B, and A biotypes, and *T. vaporariorum*. Positive reactions are indicated by the capital letter corresponding to the assay performed, and negative reactions are indicated by lower case letters. Dashes indicate that no test was performed with the specified primer set and population.

Assayed population	Species and Biotype	Primer Sets			
		Q Biotype	B Biotype	A Biotype	*T. vaporariorum*
ColonyQ	*B. tabaci* Q	E M I	E m i	- - -	- - -
Almeria16		E M -	- - -	e - -	e - -
Argentina24		E M -	E - -	e - -	e - -
Spain7		E M -	- - -	e - -	e - -
Burkinafaso2		- M -	E - -	- - -	- - -
ColonyB	*B. tabaci* B	e m -	E M I	e - -	e - -
Mexico31		- - -	E M -	- - -	- - -
Turkey19		e - -	E M -	e - -	e - -
USA55		e - -	E M -	e - -	e - -
USA72		- - -	- m -	- - -	- - -
ColonyA	*B. tabaci* A	- - -	e - -	E M I	- - -
Mexico25		e - -	e - -	E M -	e m -
EN22		e - -	e - -	E M -	e - -
BIO24		e - -	- - -	E M -	e - -
BIO13		- - -	- - -	- M -	- - -
MexicoTv	*T. vaporariorum*	e - -	e - -	e m -	E M I
Canada26		- - -	- - -	- - -	E M -
Canada288		e - -	e - -	e - -	E M -
Canada94		e - -	e - -	e - -	E M -
Canada88		- - -	- - -	- - -	- M -
Sudan1	Non-target	e - -	e - -	e - -	e - -
Cameroon7		e - -	e - -	e - -	e - -
EN3		e - -	e - -	e - -	e - -
T. abutiloneus		e - -	e - -	e - -	e - -

Table IV.2. *B. tabaci* biotype and *T. vaporariorum* primer sets designed for specific amplification of target species and/or biotype by melting temperature analysis and helicase dependent amplification assays. Bold letters in primer sequences indicate added non-complementary 5' A/T-rich overhang sequences.

Species & Biotype	Orientation	Primer Name	Primer Sequence 5'-3'	Product Size (bp)	TM (ºC)	ΔG (kcal/mol)	Alignment Accession Numbers
Bemisia tabaci Q	Forward	QhF	**CCC**TTTCACTTCAGCTACTATGATTATTGCC	114	67.3	1.0	AF342773.1, DQ133377.1, AF342775.1, AY057138.1, DQ133378.1, AF342769.1, AF342776.1, DQ302946.1
	Reverse	QhR	CAAAGGCCAAGGGGCCTGAATTTATTG		71.8	0.0	
B. tabaci B	Forward	BhF	**CGACGACGA**GTTTCTCATCTAATCAGCAGT	123	70.0	0.9	DQ133375.1, AY057123.1, AF340215.1, AF340216.1, AM180064.1, AJ550176.1, DQ133372.1, DQ133373.1, AF321927.1, AY057140.1
	Reverse	BhR	**TCGTCG**ACCTCAAACAATAAACCCTAGA		67.5	0.9	
B. tabaci A-like	Forward	AhF	**CGC**AATATTTGTGGGAGTAAACCTGACA	93	68.4	0.9	AY057122.1, AY057124.1, DQ133368.1, AY057125.1, AF342770.1, DQ133376.1, DQ133371.1, AY057128.1, AY057126.1, AF342772.1, AY057133.1, AY057127.1
	Reverse	AhR	CAATCAGGGTAGTCAGAATACCGACGAGGT		70.3	0.0	
Trialeurodes vaporariorum	Forward	TvhF	CGCGCGTTTGTCATTTAATCCCCTTAC	99	71.0	1.0	AF110708.2, JX841216.1, JX841220.1, JQ995231.1, LN614547.1, KJ475452.1, HE863766.1, JF693934.1, KF991608.1, KC843064.1
	Reverse	TvhR	GAAGAATTACCCAAAATCACCCCTGTG		67.3	0.5	

using ClustalX2 (Larkin et al., 2007), and consensus sequences for *B. tabaci* Q, B, and A-like biotypes, and *T. vaporariorum* were created using GeneDoc (Nicholas et al., 1997). *B. tabaci* biotype- and *T. vaporariorum*-specific consensus sequence alignments were used to identify regions of divergence and single nucleotide polymorphisms for type-specific primer design. Discriminatory forward and reverse primers were identified for each of the four targets, and primer stability, forward/reverse compatibility, and melting temperature were evaluated using Primer3 (Table IV.2; Untergrasser et al., 2012). Primer-target specificity was assessed using NCBI BLASTn. Primers were designed to have high melting temperatures (between 67-72ºC) and short amplicons (between 90-130-bp) for compatible conditions with the isothermal HDA reactions. Some primers (QhF, BhF, BhR, and AhF) were designed and synthesized with custom 5' overhang sequences (Afonina et al., 2007; Arif et al., 2013) added to fit within optimal melting temperature and GC content ranges while maintaining primer stability (Table IV.2). Target PCR amplicons for the *B. tabaci* Q, B, and A-like, and *T. vaporariorum* primer sets were 114-, 123-, 93-, and 99-bp, respectively. All primers were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). These primer sets will be referred to as the Q, B, A, and T sets, respectively, throughout the chapter.

End-point PCR Validation

Gradient PCRs were performed on each of the four primer sets using 10 µL of 2X GoTaq Green Master Mix (Promega, Madison, WI), 6 µL nanopure water, 1 µL of 5 µM forward primer, 1 µL of 5 µM reverse primer, and 2 µL of 20 ng/µL target DNA for a 20 µL reaction. Thermocycling conditions were 94˚C for 3 minutes; 35 cycles of 94˚C for 20 seconds, 55-70˚C for 30 seconds, and 72˚C for 40 seconds; and 72˚C for 3 minutes. Reactions were performed in a Biometra TProfessional Thermocycler (Göttingen, Germany). Gradient PCR products were separated by electrophoresis in a 1.5% agarose gel with 4 µL SYBR® Safe DNA Gel Stain

(Invitrogen, Carlsbad, CA) per 100 mL gel and a 1 Kb Plus DNA Mass Ladder (Invitrogen) in 1X TAE.

Specificities of the four primer sets were tested in three replicates of end-point PCR. Reactions were performed as described above but with an annealing temperature of 62°C. Positive control reactions consisted of the whitefly type-specific primer set and positive control target DNA (extracted from colony whiteflies) at 40 ng per reaction, while negative control reactions contained the target type-specific primer set and nanopure water (no template control; NTC). Test reaction templates consisted of crude lysate DNA extractions (DNA at variable concentrations) from the target whitefly type and three other non-target whitefly types. Three non-Q, -B, and -A biotype populations and *T. abutiloneus* were also tested to verify specificity. PCR products were separated by electrophoresis in a 1.5% agarose gel as described above.

Melting Temperature Analysis

After the target specificity of each primer set was determined, MTA assays were performed. Target amplification was carried out in reaction mixtures of 20 µL with 10 µL Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen), 0.12 µL BSA, 0.8 µL of both forward and reverse primers (5 µM each), 1 µL template DNA (variable concentrations), and 7.28 µL nanopure water. Positive control reactions consisted of control target DNA and negative control (NTC) reactions contained nanopure water. Negative test (non-target) reactions performed for the Q, B, A, and T primer sets consisted of 30 ng of B, Q, T, and A template DNA added, respectively. Cycling conditions were as described for end-point reactions followed by rapid cooling, a 2 minute hold at 25°C, and a thermal ramp from 65°C to 95°C to obtain amplicon melting. Melting was measured by the decrease in fluorescence as double stranded was denatured, causing the intercalating dye to be released and fluorescence to reduce. Assays were

performed in a Rotor-Gene 6000 thermocycler (Corbett Research, Sydney, Australia). Melt data was analyzed with Rotor-Gene 6000 Series Software 1.7 (Build 87).

Helicase Dependent Amplification

Isothermal HDA was performed using the IsoAmp® II Universal tHDA Kit (New England Biolabs, Beverly, MA). The two-step tHDA protocol was followed with slight modifications to reagent concentrations in the 50 µL reactions. Mix A contained 2.5 µL 10X Annealing buffer II, 0.8 µL of both forward and reverse primers (at 5 µM), 2 µL template DNA (50-60 ng/µL), and 18.9 µL nanopure water for 25 µL total volume. Mix A was incubated at 95°C for 2 minutes in a Biometra TProfessional Thermocycler, immediately placed on ice, briefly centrifuged, and placed again on ice. Mix B contained 2.5 µL 10X Annealing buffer, 2 µL $MgSO_4$ (100 mM), 4 µL NaCl (500 mM), 3.5 µL IsoAmp® dNTP Solution, 3.5 µL IsoAmp® Enzyme Mix, and 9.5 µL nanopure water for 25 µL total volume. Mix B was added to Mix A, mixed gently by pipetting, and reactions were incubated at 65°C for 90 minutes in a Biometra TProfessional Thermocycler. Positive and NTC controls were performed using primers with template or nanopure water provided by the HDA kit. Amplicons were visualized by electrophoresis in a 1.5% agarose gel as described above for gradient PCR products.

Results

End-point PCR

Gradient PCRs resulted in amplification of target templates from 55ºC to at least 63.1ºC for each primer set (Figure IV.1). The A set clearly amplified template DNA up to approximately 65.8ºC (lane 9), Q and B up to 64.5 ºC (lane 8), and T up to approximately 63.1ºC (Lane 7).

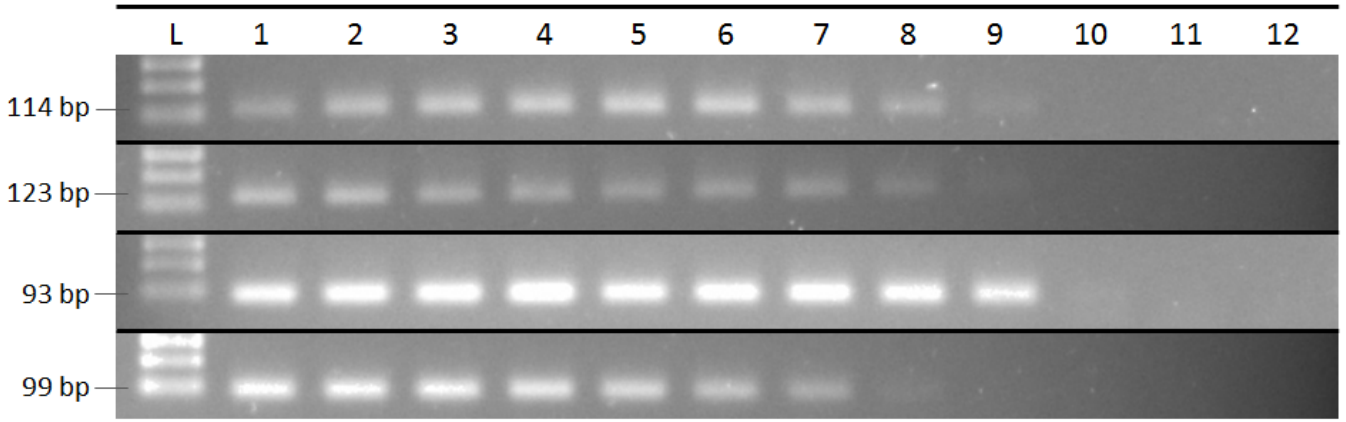


Figure IV.1. Gradient PCR amplification (annealing temperature: 55°C - 70°C) of target template DNA using designed *B. tabaci* biotype and *T. vaporariorum* primer sets. Amplification is clearly visible with each primer set from 55°C through 63°C (lane 7). Annealing temperature tests represented in each lane were the following: lane 1 - 55°C; lane 2 – 56.4°C; lane 3 – 57.7°C; lane 4 – 59.1°C; lane 5 – 60.4°C; lane 6 – 61.8°C; lane 7 – 63.1°C; lane 8 – 64.5°C; lane 9 – 65.8°C; lane 10 – 67.2°C; lane 11 – 68.6°C; lane 12 – 70.0°C.

Allowing at least 1ºC of thermocycler error, 62ºC was chosen as an optimal annealing temperature for running assays using all four primer sets concurrently.

Primer sets Q, A, and T were validated as specific to their target biotype or species by end-point PCR (Figure IV.2; Table IV.1). Each set amplified their three tested target individuals from different populations and their positive control (colony or morphologically identified at 40 ng/reaction) target only and did not amplify non-targets. The B set amplified target and control samples with high sensitivity; however, it also amplified tested Q biotypes, though with lower sensitivity (Figure IV.2; Table IV.1). The B set, therefore, could not be confirmed as target-specific in end-point assays at the tested annealing temperature of 62ºC.

Melting Temperature Analysis

Melting curve fluorescence readings were obtained for a total of 28 samples using the four primer sets. For each primer set, five target individuals (different populations) were tested with a non-target template test and a NTC. Low resolution melting curve derivative plots (-dF/dT against T) showed clear melt peaks corresponding to amplicon melting temperatures (Tm) (Figures IV.3-IV.6). Average melting temperatures (Tm), standard deviations (SD), and observed Tm ranges were recorded for each primer set test group (Table IV.3). All tested populations produced clear melting peaks with target primers except for one of the five tested B populations (USA72; Figure IV.4). Of the negative tests, the B primer set with non-target Q DNA showed a small curve in the B target observed range; this curve, however, was not interpreted as a clear melting peak because it was indistinguishable from background noise and because it fell below the determined peak threshold (Figure IV.4). All other non-target tests and all NTC were clearly negative for amplicon melting.

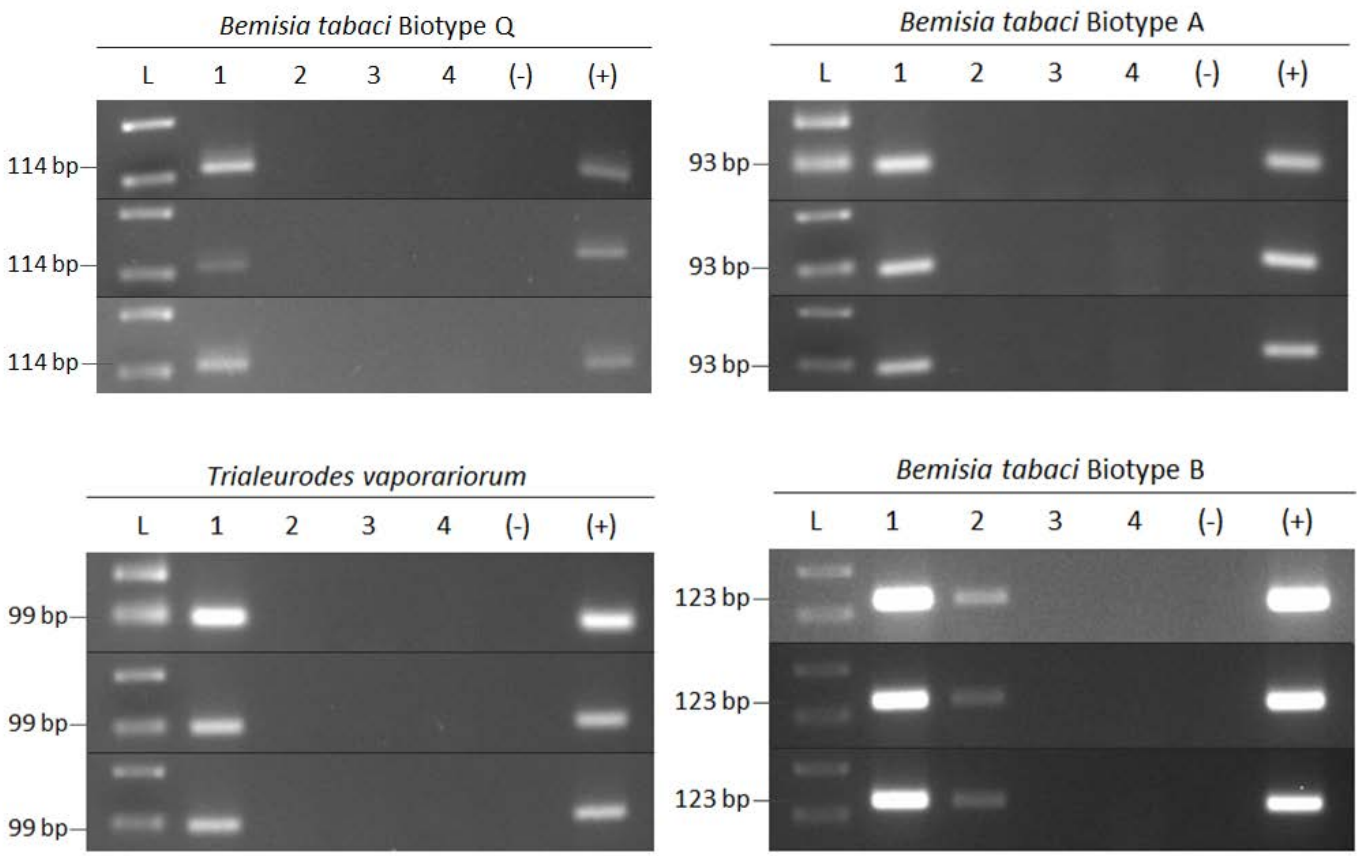


Figure IV.2. Replications of end-point *B. tabaci* biotype and *T. vaporariorum* primer sets specificity validation reactions. Lane 1 in each gel shows amplified DNA of primer targets; lanes 2-4 show the results of non-target DNA tests with only the Biotype B primers resulting in non-target Q amplification (lane 2; *Bemisia tabaci* Biotype B gel). Negative and positive controls are represented by (-) and (+), respectively.

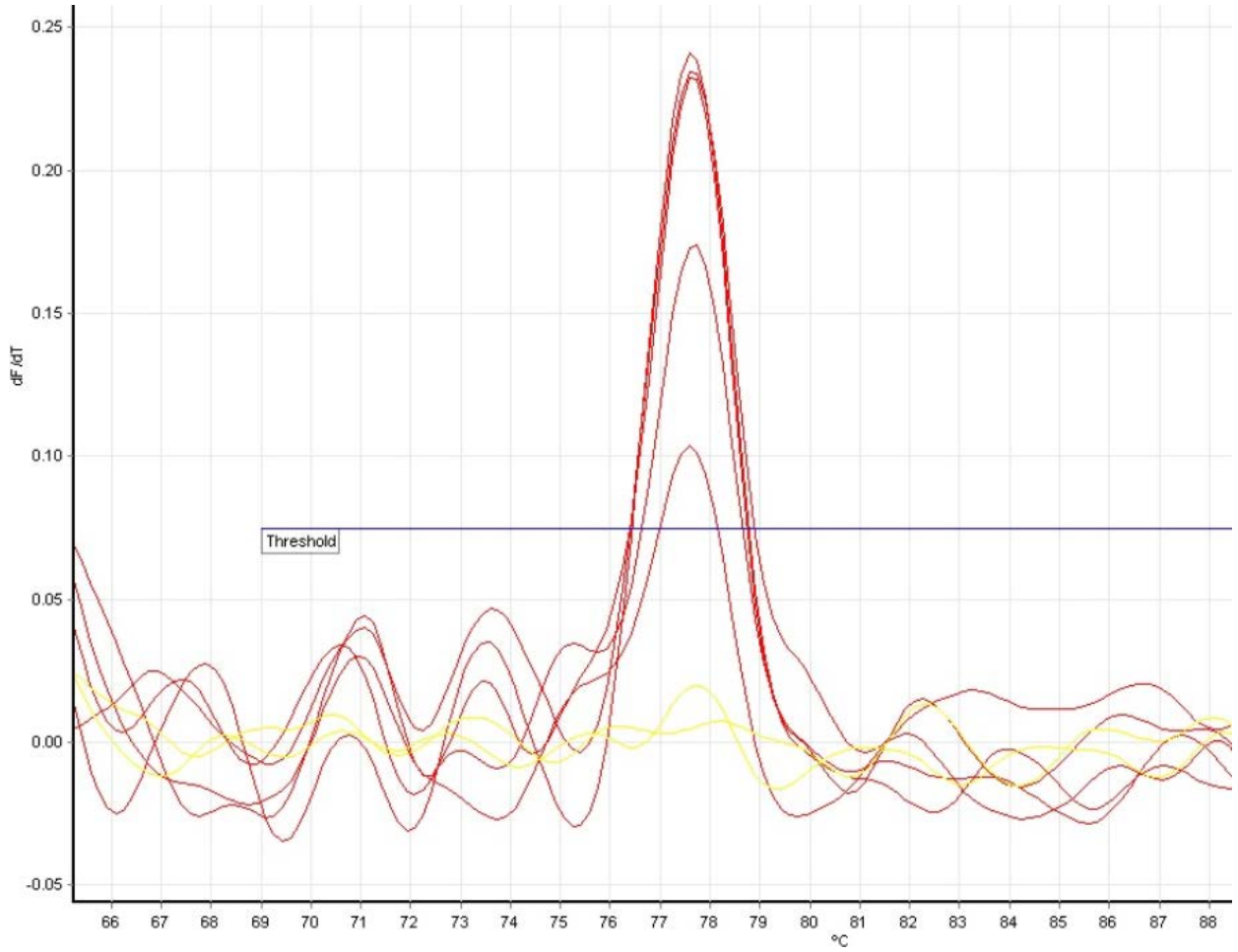


Figure IV.3. Melting curve derivative plot of populations amplified by the *B. tabaci* Q biotype primer set. Red curves represent tested Q biotype populations. Yellow curves represent the non-target test and NTC.

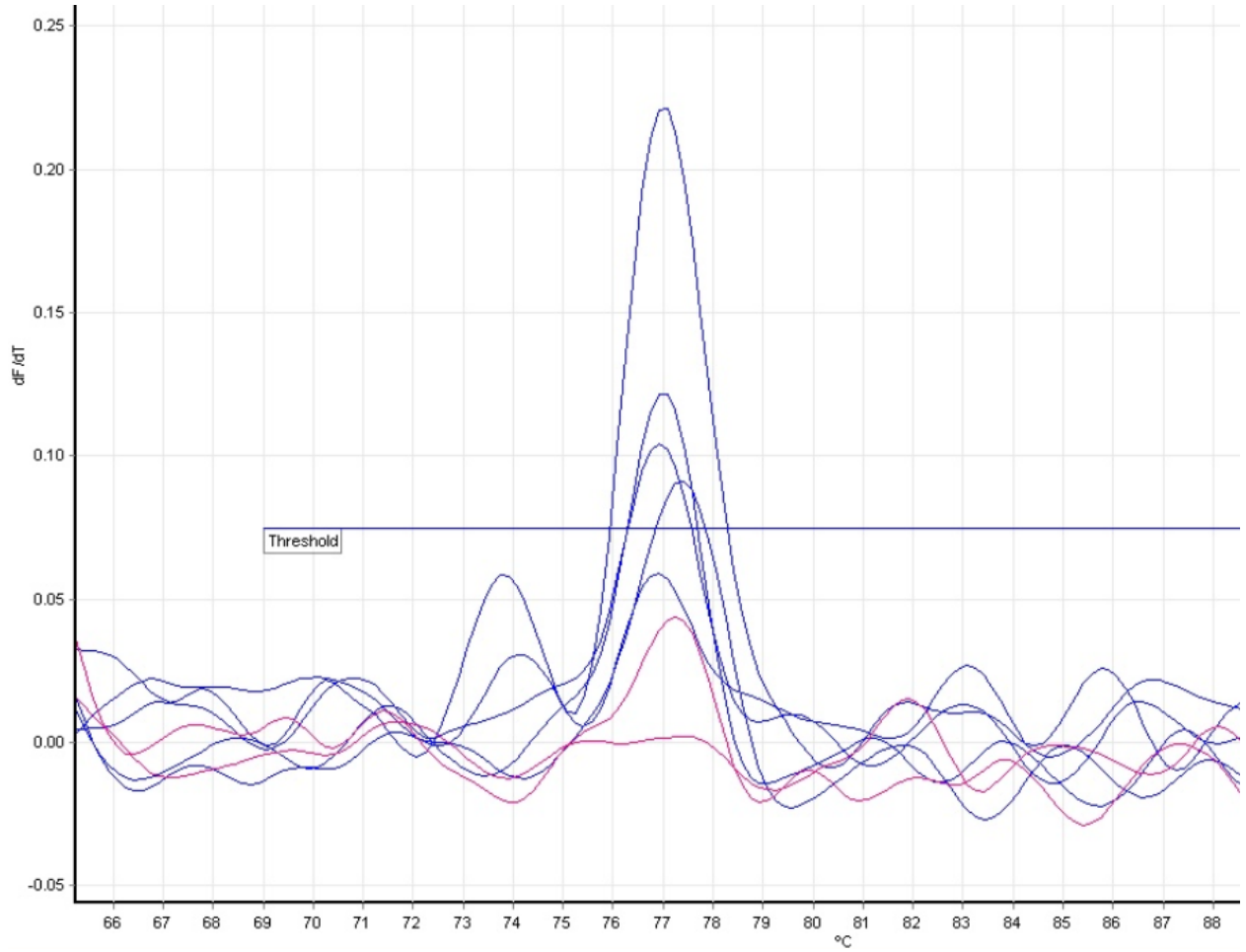


Figure IV.4. Melting curve derivative plot of populations amplified by the *B. tabaci* B biotype primer set. Blue curves represent tested B biotype populations. Pink curves represent the non-target test and NTC.

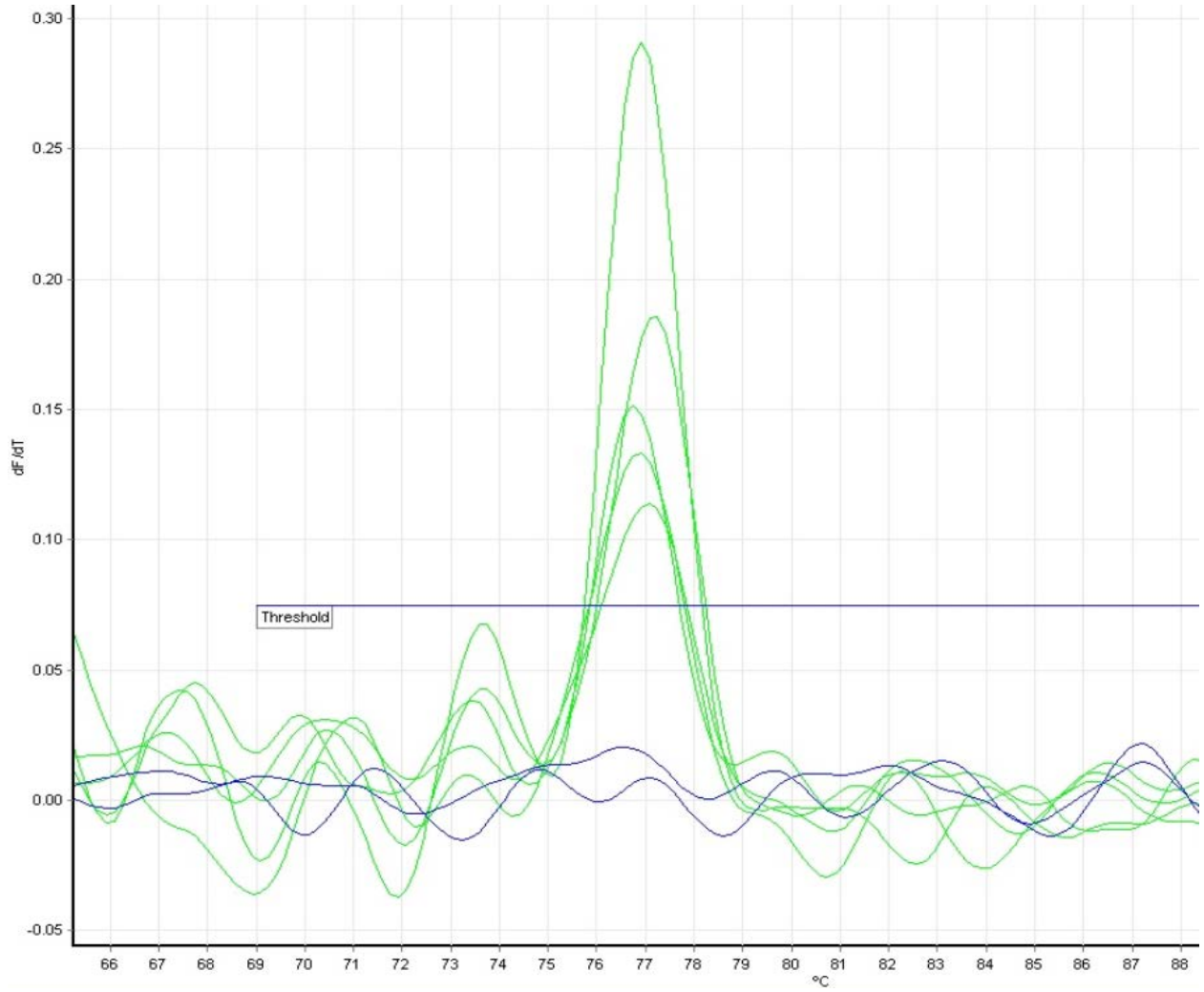


Figure IV.5. Melting curve derivative plot of populations amplified by the *B. tabaci* A biotype primer set. Green curves represent tested Q biotype populations. Blue curves represent the non-target test and NTC.

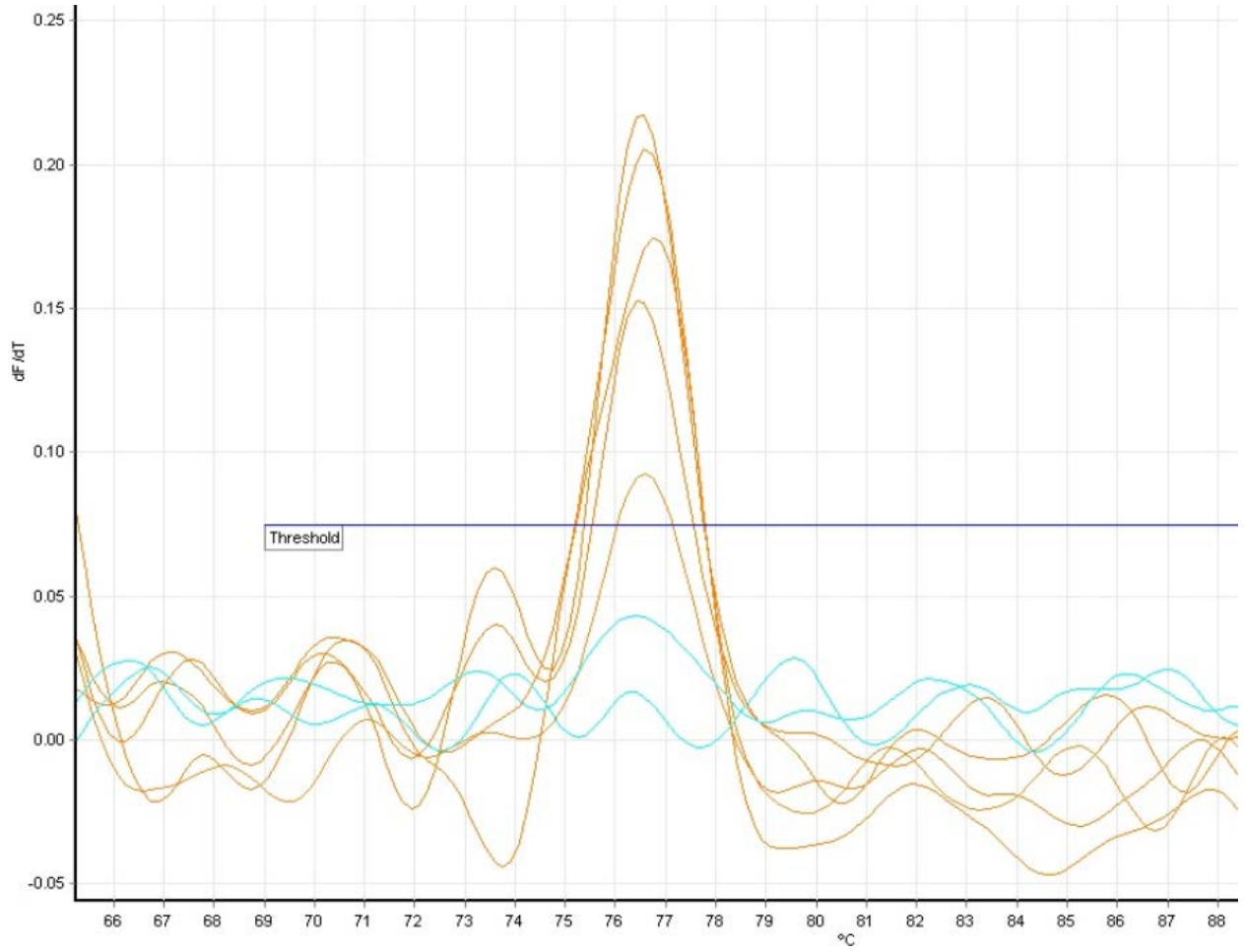


Figure IV.6. Melting curve derivative plot of populations amplified by the *T. vaporariorum* primer set. Orange curves represent tested Q biotype populations. Light blue represent the non-target test and NTC.

Table IV.3. Melting temperature analysis including average melting temperatures (Tm), standard deviations (SD), and observed Tm ranges for each *B. tabaci* biotype and *T. vaporariorum* primer set target test group.

Assayed Population	Primer Set	Tm (°C)	Mean (°C)	SD	Observed Tm Range (°C)
ColonyQ	Q	77.65	77.64	0.022	77.6-77.65
Argentina24		77.6			
BurkinaFaso2		77.65			
Almeria16		77.65			
Spain7		77.65			
ColonyB	B	76.9	77.06	0.197	76.9-77.35
Mexico31		77.0			
Turkey19		77.35			
USA55		77.0			
USA72		-			
ColonyA	A	76.85	76.95	0.170	76.75-77.15
BIO13		76.75			
BIO24		76.9			
EN22		77.15			
Mexico25		77.1			
MexicoTv	T	76.6	76.59	0.102	76.5-76.75
Canada26		76.5			
Canada288		76.5			
Canada94		76.6			
Canada88		76.75			

Pre- and post-melt regions were selected on the raw melt curve to align and normalize the curves (Leading range: 70.59 – 72.49°C; trailing range: 82.92 – 84.76°C; Figure IV.7). Normalized curves showed clustering based on the primer set used for amplification. The Q-amplified population melt curves and T- amplified population melt curves clustered separately and distinctly from others. The B- and A-amplified melt curves intermixed, clustering together but distinct from Q and T. Difference graphs were plotted against each melted control sample (Appendix 2). As with the normalized curves, difference graphs showed separation of Q and T amplicon melting and intermixed B and A melting.

Helicase Dependent Amplification

All four targets were amplified by their corresponding primer sets isothermally by HDA (Figure IV.8). Because the B set showed non-target Q amplification in end-point PCRs, the B primer set was additionally tested for amplification of control Q DNA (B/Q lane; Figure IV.8). Q was not amplified by the B set. A band appears in the B/Q lane; however, this is from primer dimers as is seen also at the same migration distance in the B lane. No band is present in the B/Q test at 123 bp, which would be expected if the B set amplified Q by HDA. Each set is specific for their target in isothermal HDA tests.

Discussion

The rapid, specific discrimination of *B. tabaci* biotypes has been an ongoing objective of whitefly researchers since the discovery of biological and genetic variation within this species in the late 1980s through early 1990s. Reliable identification and differentiation of these biotypes is critical to both *B. tabaci* research and agricultural biosecurity. As previously mentioned, a number of assays have been developed for the discrimination of the Q, B, and A biotypes, which

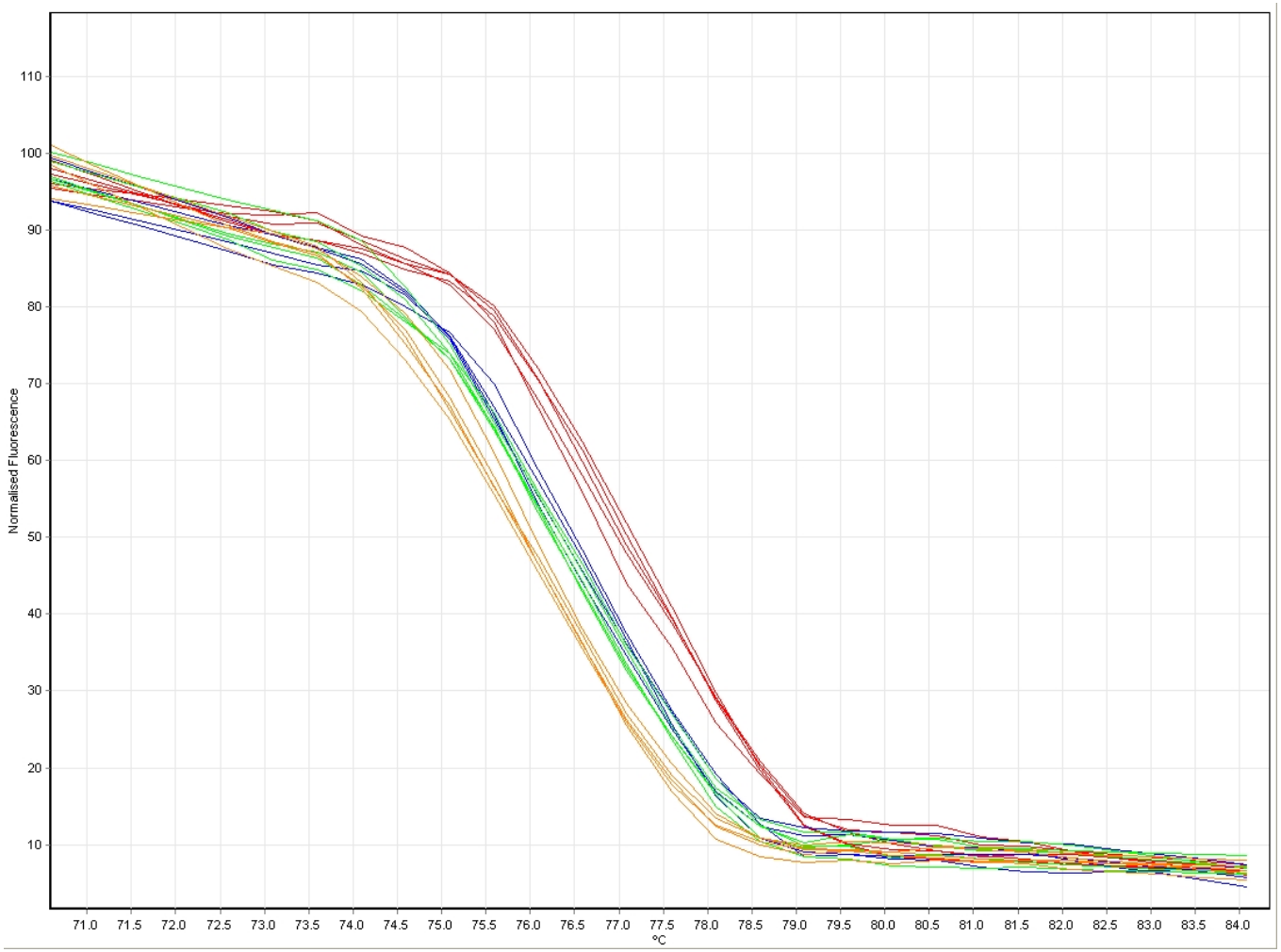


Figure IV.7. Normalized florescence curves for *B. tabaci* biotype and *T. vaporariorum* populations amplified by designed primer sets. Red curves represent *B. tabaci* Q biotype population samples; blue represent B; green represent A; and orange represent *T. vaporariorum* samples.

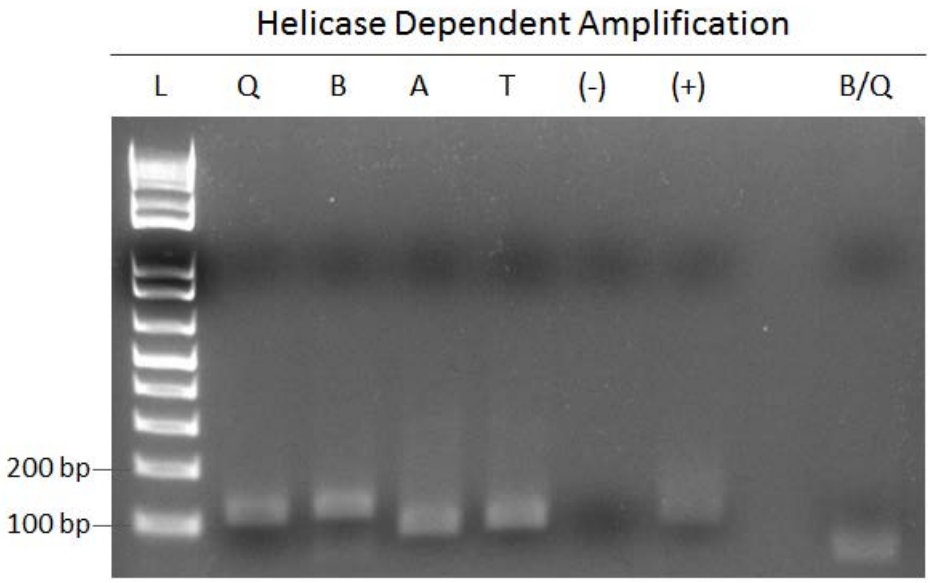


Figure IV.8. Isothermal helicase dependent amplification of *B. tabaci* biotypes and *T.vaporariorum* using designed type-specific primer sets. Lanes Q, B, A, and T contain isothermally amplified target DNA; Lane B/Q contains a band resulting from primer dimer formation, not amplification, by B primers tested with non-target Q template DNA.

have been of the most interest within the U.S. The development of the first MTA and HDA assays for detection of *B. tabaci* biotypes Q, B, and A, and *T. vaporariorum* is reported herein.

Four primer sets for use in end-point PCR, MTA or other qPCR applications, and isothermal HDA assays were designed to be highly specific to their target whitefly type. The Q, A, and T primer sets were validated for specificity within end-point PCRs and showed specific amplification in MTA and HDA assays. The B primer set resulted in target amplification in all three assays but it could not be concluded as target-specific in end-point reactions as it also amplified Q DNA though with lesser sensitivity. Differentiation of Q and B is not an uncommon challenge, as these two biotypes show considerable homology within the mtCOI gene. This set, amplified Q DNA though with lesser sensitivity. Differentiation of Q and B is not an uncommon challenge, as these two biotypes show considerable homology within the mtCOI gene. This set, however, did show specific amplification in MTA and HDA. Despite the lack of a high degree of specificity in the B primer set, the concurrent use of all four sets (or only the Q, B, and A sets if morphologically identified to *B. tabaci*) in any of the three reported assays provides reliable identification of these three biotypes or *T. vaporariorum*.

In the MTA assays, variability observed in the results of tested populations for each primer set was likely due to different DNA concentrations added to individual reactions and/or nucleotide content differences among amplified biotypes. In the derivative plots, measured peak dF/dT varied among tested populations within a biotype (Figures IV.3-IV.6). These differences likely corresponded to the amount of template DNA added to the reaction. Using crude extracts having variable DNA concentrations was relevant to our objective of creating an effective, rapid identification tool for which the DNA concentration extracted from an individual whitefly need not be modified prior to assay, minimizing the time required for identification of the whitefly species and biotype. Sensitivities of the primers were not tested; however, all tested sample DNA

was extracted from individual whiteflies, and all reactions but one showed strong peaks above the designated threshold, thereby demonstrating the practicality of this approach. Lower peak dF/dT could reflect lesser amplification resulting from low primer specificity/sensitivity for a tested template genotype. The curve observed for the non-target Q test with B primers could be a product of this occurrence (Figure IV.4). Though the curve was not called a peak because it did not meet the threshold, florescence measured from some melting of a lower number of amplicons likely occurred resulting in the raised curve within the B biotype 76.9-77.35 Tm range. Therefore, either genotype or nucleotide content differences affects melting curve results in both peak dF/dT and observed Tm. Achieving highly consistent melting temperatures within *B. tabaci* biotypes is difficult, if not impossible, because there is genetic variation within biotypes. Different haplotypes, genetic variants of biotypes, and even population level SNPs are common (Frohlich et al., 1999). As such, in practice, observing a Tm within a specific range designated for calling a particular biotype is more realistic than expecting a specific Tm per biotype. Performed in its current state, this MTA assay can detect *B. tabaci* Q, B, A biotype, or *T. vaporariorum* by running a sample with the four primer sets.

HDA is a desirable alternative to traditional PCR because it can be performed isothermally without need for a thermocycler and is therefore deployable in field setting identification. The four primer sets designed herein successfully amplified their corresponding target at 65°C using HDA chemistry. The question of specificity for the one end-point unspecific B primer set was tested. This reaction showed that in HDA the B primer set does not amplify the homologous Q biotype. These four primer sets can be applied for whitefly species and biotype discrimination in different amplification assay strategies, thus expanding the *B. tabaci* identification protocol repertoire.

REFERENCES

Abdullahi, I., Atiri, G.I., Thottappilly, G., and Winter, S. 2004. Discrimination of cassava-associated *Bemisia tabaci* in Africa from polyphagous populations, by PCR–RFLP of the internal transcribed spacer regions of ribosomal DNA. Journal of applied entomology 128(2):81-87.

Adachi, T. Umezawa, T. Yamaguchi, H., Kitamura, T. Honda, K. Shibao, M., Nasu, Y., and Tanaka, H. 2010. Identification of *Bemisia tabaci* B and Q biotypes in Osaka Prefecture by DNA sequencing and a LAMP method. Annual Report of the Kansai Plant Protection Society 52:103-104.

Barbosa, L.F., Marubayashi, J.M., De Marchi, B.R., Yuki, V.A., Pavan, M.A., Moriones, E., Navas-Castillo, J. and Krause-Sakatea, R. 2014. Indigenous American species of the *Bemisia tabaci* complex are still widespread in the Americas. Pest Management Science 70:1440–1445.

Brown, J. K. 1994. The Status of *Bemisia tabaci* Genn. as a pest and vector in world agroecosystems. FAO Plant Protection Bulletin 42:3-32.

Costa, H.S. and Brown, J.K. 1991. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction.Entomologia Experimentalix et Applicata 61:211–19.

Cervera, M.T., Cabezas, J.A., Simon, B., Martinez-Zapater, J.M., Beitia, F., Cenis, J.L. 2000. Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on AFLP analysis. Bulletin of Entomological Research 90:391–96.

Dickey, A.M., Osborne, L.S., Shatters, R.G., and McKenzie, C.L. 2013. Identification of the MEAM1 cryptic species of *Bemisia tabaci* (Hemiptera: Aleyrodidae) by loop-mediated isothermal amplification. Florida Entomologist 96(3):756-64.

De Barro, P.J. 2005. Genetic structure of the whitefly *Bemisia tabaci* in the Asia-Pacific region revealed using microsatellite markers. Molecular Ecology 14:3695–718.

De Barro, P.J., Liu, S.S., Boykin, L.M., and Dinsdale, A.B. 2011. *Bemisia tabaci*: a statement of species status. Annual Review of Entomology 56:1–19.

Esterhuizen, L.L., Mabasa K.G., van Heerden, S.W., Czosnek, H., Brown, J. K., van Heerden, H., and Rey, M.E.C. 2013. Genetic identification of members of the *Bemisia*

tabaci cryptic species complex from South Africa reveals native and introduced haplotypes. Journal of Applied Entomology 137:122–135.

Frohlich, D.R., Torres-Jerez, I., Bedford, I.D., Markham, P.G., and Brown, J.K. 1999. A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. Molecular Ecology 8:1683–91.

Gawel, N.J., Bartlett, A.C. 1993. Characterization of differences between whiteflies using RAPD-PCR. Insect Molecular Biology 2:33–38.

Gennadius, P. 1889. Disease of tobacco plantations in the Trikonia. The aleurodid of tobacco. Ellenike Georgia 5:1-3.

Hsieh, C.H., Wang, H.Y., Chen, Y.F., and Ko, C.C. 2012. Loop-mediated isothermal amplification for rapid identification of biotypes B and Q of the globally invasive pest, *Bemisia tabaci*, and studying population dynamics. Pest Management Science 68:1206-13.

Liu, S.S., Colvin, J., and De Barro, P.J. 2012. Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? Journal of Integrative Agriculture 11:176–186.

Ma, D.Y., Li, X.C., Dennehy, T.J., Lei, C.L., Wang, M., Degain, B.A., and Nichols, R.L. 2009. Utility of mtCO1 polymerase chain reaction-restriction fragment length polymorphism in differentiating between Q and B whitefly *Bemisia tabaci* biotypes. Insect Science and its Application 16:107–14.

Mound, L.A., and Halsey, S.H. 1978. Whiteflies of the World. New York: Wiley. 340 pp.

Papayiannis, L.C., Brown, J.K., Seraphides, N.A., Hadjistylli, M., Ioannou, N., Katis, N.I. 2009. A real-time PCR assay to differentiate the B and Q biotypes of the *Bemisia tabaci* complex in Cyprus. Bulletin of Entomology Research 99(6):573-82.

Perring, T.M., Cooper, A., Kazmer, D.J. 1992. Identification of the poinsettia strain of *Bemisia tabaci* (Homoptera: Aleyrodidae) on broccoli by electrophoresis. Journal of Economic Entomology 85:1278–84.

Quaintance, A.L. 1900. Contribution towards a monograph of the American Aleurodidae. US Department of Agriculture, Technical Services, Bureau of Entomology 8:9-64.

Shatters, R.G. Jr, Powell, C.A., Boykin, L.M., He, L.S., McKenzie, C.L. 2009. Improved DNA barcoding method for *Bemisia tabaci* and related Aleyrodidae: development of universal and *Bemisia tabaci* biotype-specific mitochondrial cytochrome c oxidase chain reaction primers. Journal of Economic Entomology 102:750–58.

CHAPTER V

APPLICATION OF E-PROBE DIAGNOSTIC NUCLEIC ACID ASSAY (EDNA) TO PLANT PATHOGENS WITHIN INSECT VECTOR NEXT GENERATION SEQUENCING DATASETS: *SPIROPLASMA KUNKELII* DETECTION IN GRAY LAWN LEAFHOPPER, *EXITIANUS EXITIOSUS*, TRANSCRIPTOMES

Abstract

E-probe diagnostic nucleic acid assay (EDNA) is a novel bioinformatic pipeline for the detection and identification of pathogens within next generation sequencing (NGS) datasets without the need for prior sequence data assembly and analysis. This diagnostic tool has been applied to and validated for the detection of pathogens (plant viruses, bacteria, mollicutes, fungi, oomycetes, and human pathogens) on and within plants. As many plant pathogens are disseminated via insects, the objective of this project was to test and validate the EDNA pipeline for sensitive and specific pathogen detection in an insect vector. *Spiroplasma kunkelii*-specific electronic probes (e-probes) were designed based on sequences identified as unique to *S. kunkelii* by comparing the species' genome to that of the close phylogenetic relative *S. citri*. These e-probes were filtered for target specificity by query against GenBank. The filtered *S. kunkelii*-specific e-probes were then tested *in silico* by generating mock sample databases (MSDs) with different abundance levels of the target pathogen (high, medium, low, very low)

and insect host sequences. After theoretical use of the designed e-probes and EDNA pipeline for this system was shown to be effective, validation tests using Illumina NextSeq500 raw sequence datasets derived from *S.kunkelii*-infected versus naïve *Exitianus exitiosus* (Uhl.) (Hemiptera; Cicadellidae) transcriptomes were conducted. *S. kunkelii* e-probes were able to significantly detect the pathogen sequences at abundances as low as in the tenths of a percentage in the MSDs as well as in the *S. kunkelii*-infected *E. exitiosus* sequence dataset. Insect-derived NGS datasets can be rapidly queried for pathogens of interest without any prior traditional read processing or assembly with the use of EDNA.

Introduction

As biological research is experiencing an era of genomic revolution, sequencing technologies that are currently considered technologically advanced are likely to become streamlined and user-friendly over time. Next generation sequencing (NGS) will likely become simplified to the point where producing metagenomic data will be as conventional as running a PCR assay. Until that time, working on other aspects of NGS, such as the bioinformatic processing of its produced data, will facilitate the quick adaptation of the technology to rapid detection purposes. While the in-house computational demand remains one of the central challenges of NGS, analysis of NGS datasets usually presents the additional challenge of requiring lengthy data assembly and annotation to draw any conclusions. In laboratories that have the resources required for producing and storing vast amounts of sequencing data, such as government labs, NGS data would be extremely useful for a number of research purposes. In this work, our goal was to address the possible application of NGS for rapid diagnostic and agricultural biosecurity purposes at ports of entry. That is, agricultural products entering the

country could be screened for any and all pathogens of interest using NGS, and the results could facilitate rapid agriculture protection-related decisions.

The e-probe diagnostic nucleic acid assay (EDNA) bioinformatic pipeline was developed to rapidly screen samples for targeted pathogens in large NGS datasets (Stobbe et al., 2013). EDNA is designed for the rapid detection of targets of interest using electronic probes (e-probes) in raw genomic or transcriptomic NGS data without the need for time consuming assembly or annotation. E-probes have been designed and tested using mock sample databases (MSDs) for phytopathogenic fungi, oomycetes, bacteria, mollicutes, and viruses as well as human pathogens that could be contaminating plants (Stobbe et al., 2013; Blagden et al., 2013). To date, the validation of EDNA detection in infected plant sample NGS datasets has been tested in each of these systems and published for viruses (Stobbe et al., 2014), fungi and oomycetes (Espindola et al., 2015), and human pathogens on plants (Blagden et al., 2016) with other systems' validations in preparation for submission (Daniels et al., in preparation).

Plants can be a significant source of exotic pathogens entering the country; however, insect vectors provide yet another important means of plant pathogen introduction. As phytopathogenic mollicutes, spiroplasmas and phytoplasmas, can be disseminated among plants by insect host intermediates, the ability to quickly detect mollicutes carried by insect vectors could prompt interventions that minimize the likelihood of their spread between countries, regions, or even fields. The purpose of this study was to apply and adapt EDNA to pathogen detection in insect-derived metagenomic datasets. The ubiquitous gray lawn leafhopper, *E. exitiosus*, a vector of the corn stunt spiroplasma, *S. kunkelii*, was used as a model for insect transmitted plant pathogen detection in NGS datasets using the EDNA bioinformatic pipeline. This system was chosen for three reasons: first, *S. kunkelii* is replicative within the vector and is thereby likely to reach high numbers, providing a high chance of being sequenced and subsequently detected using EDNA; second, the full chromosomal genome of *S. kunkelii* is

available in GenBank, therefore offering an ideal template for e-probe design; and third, this system represents EDNA-based detection of a mollicute within a leafhopper NGS dataset and will be combined for future publication with detection of a walled-bacterium within a psyllid and detection of a virus within an aphid, thereby representing the validation of EDNA-based detection on a range of different pathogen classes transmitted by different vector families.

Materials and Methods

E-probe Design

S. kunkelii-specific e-probes were designed using a modified version of the Tool for Oligonucleotide Fingerprint Identification (TOFI) (Satya et al., 2008). Using this version of TOFI the target genome *S. kunkelii* (generously provided by Dr. Robert Davis from the USDA ARS, Beltsville, MD; now available in NCBI Genbank: reference sequence NZ_CP010899.1), was compared with the genome sequences available for the close relative *S. citri* (Genbank) to identify regions of sequence divergence for electronic oligonucleotide selection. The modified TOFI then restricted and optimized resulting draft e-probe lengths and removed those with homo-oligomers (five or more consecutive identical nucleotides) because NGS platforms can have trouble calling identical nucleotides that occur successively.

Draft e-probes were queried against the NCBI nucleotide database at an E-value of 1×10^{-9} to further ensure e-probe specificity to *S. kunkelii*. Decoy e-probes, which are simply the reverse sequence of the final e-probe set, were then designed to query MSDs and NGS datasets to control for false positive or random matching of e-probes within the datasets and to statistically analyze match results.

Mock Sample Database Generation and Query

To test the accuracy and estimate the sensitivity of the final *S. kunkelii* e-probe set *in silico*, simulated NGS datasets were constructed with known percentages of target (*S. kunkelii*) and background (leafhopper) sequences. MSDs were constructed using Metasim, a next generation sequencing run simulator (Richter et al., 2008). MSDs were designed by importing the *S. kunkelii* genome sequence used for e-probe design into MetaSim as the target and importing a large *Homalodisca vitripennis* (Germar) (Hemiptera; Cicadellidae) (glassy-winged sharpshooter) scaffold sequence (Accession KK961494.1) as the background/host. *H. vitripennis* was chosen to represent *E. exitiosus* background sequences because the two species are closely related, and, of the leafhoppers present in Genbank, *H. vitripennis* has the most complete genome sequence data. High (H:15-25%), medium (M:5-15%), low (L:0.5-5%), and very low (VL:0.01-0.5%) target percentage MSDs were constructed by editing the taxon profile values for the target sequence.

Each MSD was formatted and queried with the *S. kunkelii*-specific e-probes set and decoy e-probe set using the BLASTn algorithm with an E-value threshold of 50 as performed in Stobbe et al. (2013). Setting a high E-value threshold provided abundant BLASTn results which subsequently could be parsed (filtered) for analysis at different and lower (more stringent) E-values. E-probes that hybridized with their respective sequences within the dataset were termed matches. The number of matches that occurred was limited to the number of e-probes used as queries such that a specific e-probe could have only one match. A particular e-probe could, however, find its sequence within the dataset multiple times; these were termed hits. The resulting BLASTn output was then parsed at E-values of $10^{-9,}$ 10^{-6}, and 10^{-3} to report results at varying levels of stringency. These reports were used to choose an E-value at which to parse the validation results for minimal false matches or hits. Resulting matches and hits were reported for each e-probe and decoy e-probe at the three filter E-values.

Statistical Analysis

Pearson goodness of fit statistic (chi-square) with one degree of freedom and a 5% significance level was conducted on resulting e-probe matches versus decoy e-probe matches to assess their statistical difference and provide support for either detection or lack of detection with the target e-probes (Blagden et al., 2016). Chi-square statistics were then converted to p-values (http://www.socscistatistics.com/pvalues/chidistribution.aspx).

S. kunkelii Acquisition by *E. exitiosus*

Late (third-fourth) instar *E. exitiosus* were collected from bermudagrass in late September, 2014 in Stillwater, OK. Leafhoppers were placed in groups of 10 in feeding sachets with 500 µL of D10 medium (Alivizatos, 1982) with 1% fetal bovine serum (FBS) (negative control; *S. kunkelii*-naive) or 500 µL of D10 medium with resuspended *S. kunkelii* (*S. kunkelii*-infected). Feeding sachets consisted of a 1 oz. medicine cup covered with two stretched parafilm membranes with D10 medium or D10/*S. kunkelii* between the layers. To prepare feeding sachets, *S. kunkelii* CR2-3X cultures were grown to log phase at 29˚C in LD8A3 medium (Maramorosch and Harris, 1979) and evaluated by a dark-field microscope for quality. Cultures were centrifuged at 10,000 rpm for 1.5 min to form a pellet and medium supernatant was discarded. *S. kunkelii* cells were then resuspended in D10 medium at approximately 10^8 CFU/mL for feeding sachet acquisition tests. After a 24 hr acquisition access period (AAP), leafhoppers were transferred to healthy corn (*Zea mays* L.) grown to the 2-4 leaf stage for a latent period (LP) of 25 days to allow enough time for *S. kunkelii* to be acquired and begin to multiply in the vector (Nault, 1980). Corn and leafhoppers were maintained in a growth chamber with a 16 hr photoperiod at 27°C (day) and 18°C (night).

Surviving *E. exitiosus* leafhoppers were tested by PCR for *S. kunkelii* acquisition and to confirm *S. kunkelii* absence in the control group. Leafhoppers were collected from corn, chilled at -20°C for 1 min, and the right metathoracic leg was excised. Legs were placed individually in 180 µL sterilized phosphate buffered saline (1x PBS; pH 7.4) and homogenized with a plastic pestle, and DNA was extracted using the Qiagen DNeasy® Blood and Tissue Kit (Valencia, CA) following the manufacturer's instructions. *S. kunkelii* was then detected by endpoint PCR using *S. kunkelii*-specific primers F1 and R1, which amplify a segment of the spiralin gene (Barros et al., 2001). PCR reactions consisted of 10 µL GoTaq® Green Master Mix (Promega, Madison, WI), 1 µL 5 mM F1, 1 µL 5 mM R1, 2 µL template DNA (Qiagen extraction; total DNA concentration not measured), and 6 µL nanopure water. Cycling conditions used were 95°C for 3 min; 35 cycles of 95°C for 30 sec, 60°C for 1 min, 72°C for 45 sec; and 72°C for 5 min. Amplicons were visualized by gel electrophoresis in a 1.5 % agarose gel in 1XTAE with 3 µL SYBR® Safe DNA Gel Stain (Invitrogen, Carlsbad, CA) per 100 mL gel.

RNA Extraction, Sequencing, and Transcriptome Query

Total RNA was extracted from a pool of ten (five male and five female) *E. exitiosus* for both the *S. kunkelii*-naïve and *S. kunkelii*-infected groups using Trizol and Purelink™ RNA Mini Kit (Invitrogen) per the manufacturer's instructions. Resulting RNA was cleaned with Deoxyribonuclease I (amplification grade; Invitrogen) treatment to remove contaminant DNA and quantified three times using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE) before being shipped on dry ice to Cofactor Genomics (4044 Clayton Ave. Saint Louis, MO) for sequencing.

RNA was processed for library construction by Cofactor Genomics. Briefly, total RNA (no selection) was sheared to appropriate size for cDNA synthesis. Double-stranded cDNA was end-repaired and A-tailed to prepare for adaptor ligation. Indexed adaptors were ligated to sample

cDNA, and the adaptor-ligated cDNA was then size-selected on a 2% SizeSelect™ E-Gel (Invitrogen) and amplified by PCR. Library size and quality was assessed on an Agilent Bioanalyzer and library yield was quantified by qPCR using the Kapa Biosystems library quantification kit (Wilmington, MA) prior to sequencing on the Illumina NextSeq500 (San Diego, CA) as 1x75 base reads following manufacturer's protocols.

Raw transcriptome sequence datasets (*S. kunkelii*-infected and *S. kunkelii*-naive) were converted from Fastq to Fasta and formatted into a BLAST database (NCBI makeblastdb application; http://www.ncbi.nlm.nih.gov/books/NBK279688/) for query using the EDNA pipeline. The datasets were queried with *S. kunkelii* and decoy e-probe sets using the BLASTn algorithm at an E-value of 50. The results were parsed at E-values of $10^{-9,}$ 10^{-6}, and 10^{-3}. Chi-square tests were performed on the matches of e-probes and decoy e-probes at each E-value and converted to p-values.

Bioinformatic Analysis

To validate further the detection of *S. kunkelii* sequences within the *S. kunkelii*-infected transcriptome, common bioinformatic analyses were performed on both transcriptome datasets. Reads were mapped and aligned to the *S. kunkelii* chromosomal (Reference Sequence NZ_CP010899.1) and three associated plasmid (pSKU226: NZ_CP012423.1; pSKU205: NZ_CP012424.1; and pSKU76: NZ_CP012425.1) reference genomes available in Genbank using Read Mapper (Geneious 9.1.2; Auckland, New Zealand; Kearse et al., 2012) at the software-recommended medium-low sensitivity setting with five iterations and Bowtie 2 2.2.6 (Langmead and Salzberg, 2012) with the highest sensitivity setting. BLAST 2.2.27+ was performed on reads mapping to the *S. kunkelii* genomes (Altschul et al., 1990). MEGAN5 was used to determine taxonomic profiles of the datasets (El Hadidi et al., 2013).

Results

E-probe Design and MSD Query

A total of 5,264 *S. kunkelii*-specific e-probes were designed. Lengths of e-probes were set at 100-, 80-, 60-, 40-, and 25-bases to increase possible matches (compared to potential match quantities using a set of e-probes at only one length) and provide a range of lengths allowing for hits supported at different E-values. BLAST E-values are partially dependent on pairwise query-match length and queried database size. As queried database size decreases, resulting match E-values increase, potentially resulting in filtering out of positive matches that happen to have E-values above the designated threshold. Therefore, having a mix of different e-probe lengths (often generated from the same genomic loci) increases the chances of highly-supported detection of a particular target sequence compared to having e-probes of only one length. Totals of 45 100-base probes, 118 80-base probes, 373 60-base probes, 1,531 40-base probes, and 3,197 25-base probes were designed. 5,264 control decoy e-probes (reverse e-probe sequence) also were made. Resulting MSDs had the following target sequence percentages (Table V.1):

Table V.1. Percentages of target (*Spiroplasma kunkelii*) sequences present within mock sample databases constructed for testing *S. kunkelii*-specific e-probes using the EDNA pipeline.

Mock Sample Database (MSD)	High (H)	Medium (M)	Low (L)	Very Low (VL)
1	16.21%	12.71%	3.51%	0.24%
2	20.81%	10.21%	1.99%	0.01%
3	23.51%	7.06%	0.87%	0.42%

Each MSD was queried with target e-probes and decoy e-probes using the EDNA pipeline, and resulting matches, hits, decoy matches, and decoy hits were recorded (Appendix 3). Chi-square tests were performed to assess the difference between e-probe and decoy e-probe matches. Chi-square results were converted to p-values and reported for each MSD at each of the three parser E-values. Detection of target sequences by the e-probes, as assessed by the significant difference between the number of e-probe and decoy e-probe matches, was significant for each MSD, except for MSD VL-2 (0.01%), at each parser E-value (Appendix 3). Average matches, hits, decoy matches, and decoy hits were also taken for each of the four MSD target levels (H, M, L, and VL) at the three parser E-values, and chi-square and p-values were calculated and reported (Figures V.1-V.3; values recorded in Appendix 4).

The MSD designed with 0.01% target sequences (MSD VL-2) within the dataset had the lowest target percentage tested. Parsing this dataset alone at $10^{-9,}$ 10^{-6}, and 10^{-3} resulted in no matches at 10^{-9} and insignificant levels of detection at 10^{-6} and 10^{-3} (Table V.2; Appendix 3). Considering the lack of significant detection in MSD VL-2 and the significant detection found for all other MSDs individually (Appendix 3) or when taking the averages for each MSD level (Appendix 4), the designed e-probes were shown to detect *S. kunkelii* sequences at a sensitivity level of at least 0.22% target (VL Average; Appendix 4) in the dataset. The detection limit as estimated by MSD testing and chi-square analysis could therefore be assumed to lie between 0.22% and 0.01% target sequences in a queried dataset.

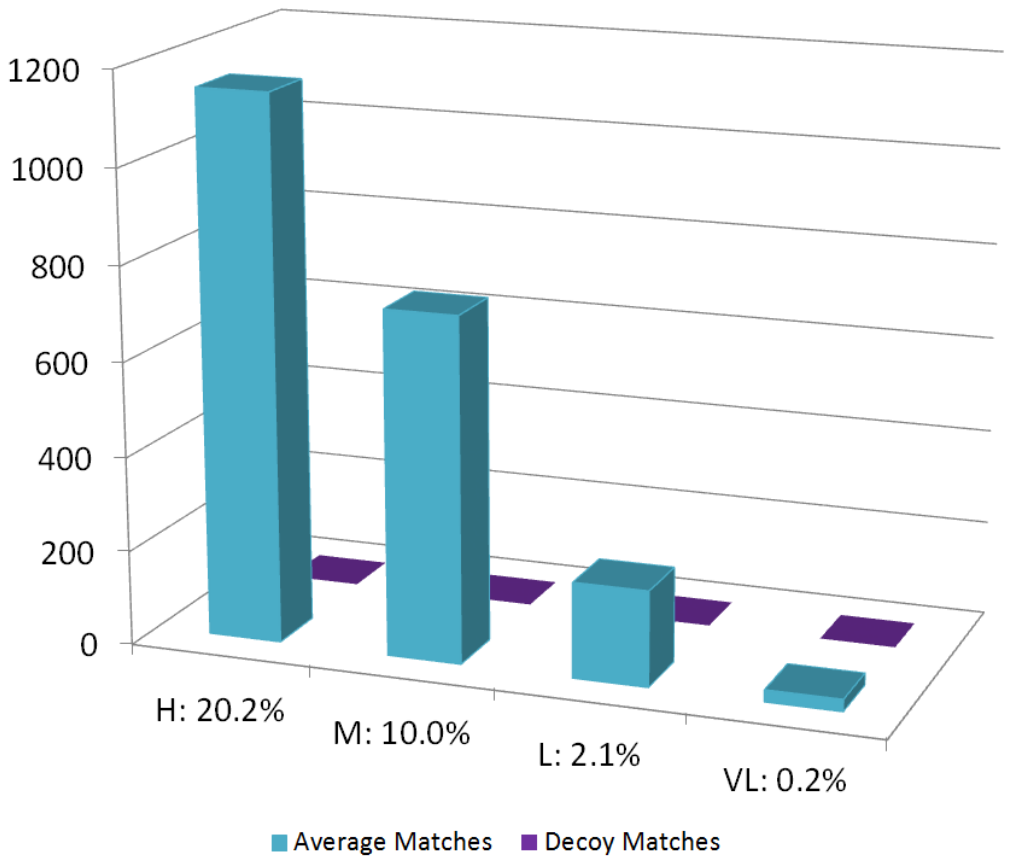


Figure V.1. Average matches in high (H), medium (M), low (L), and very low (VL) target (*Spiroplasma kunkelii*) percentage mock sample databases (MSDs) queried with *S. kunkelii*-specific e-probes and decoy e-probes using the EDNA pipeline and parsed at 10^{-9}.

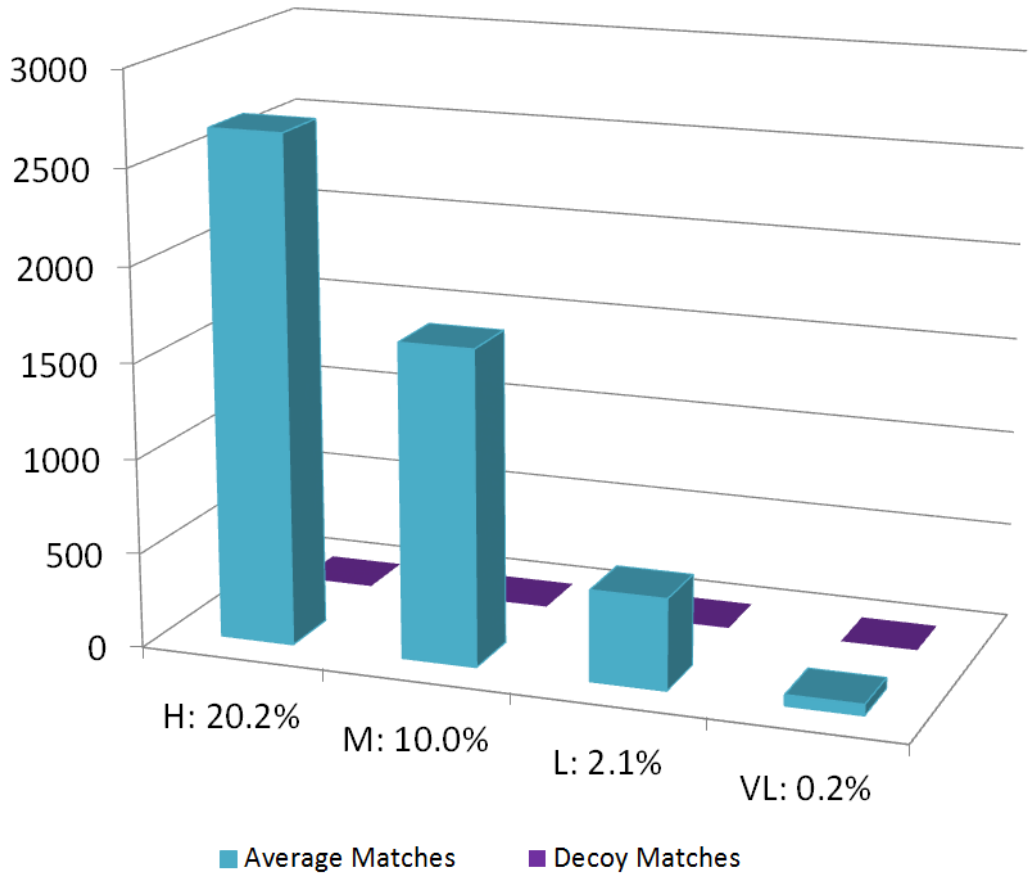


Figure V.2. Average matches for high (H), medium (M), low (L), and very low (VL) target (*Spiroplasma kunkelii*) percentage mock sample databases (MSDs) queried with *S. kunkelii*-specific e-probes and decoy e-probes using the EDNA pipeline and parsed at 10^{-6}.

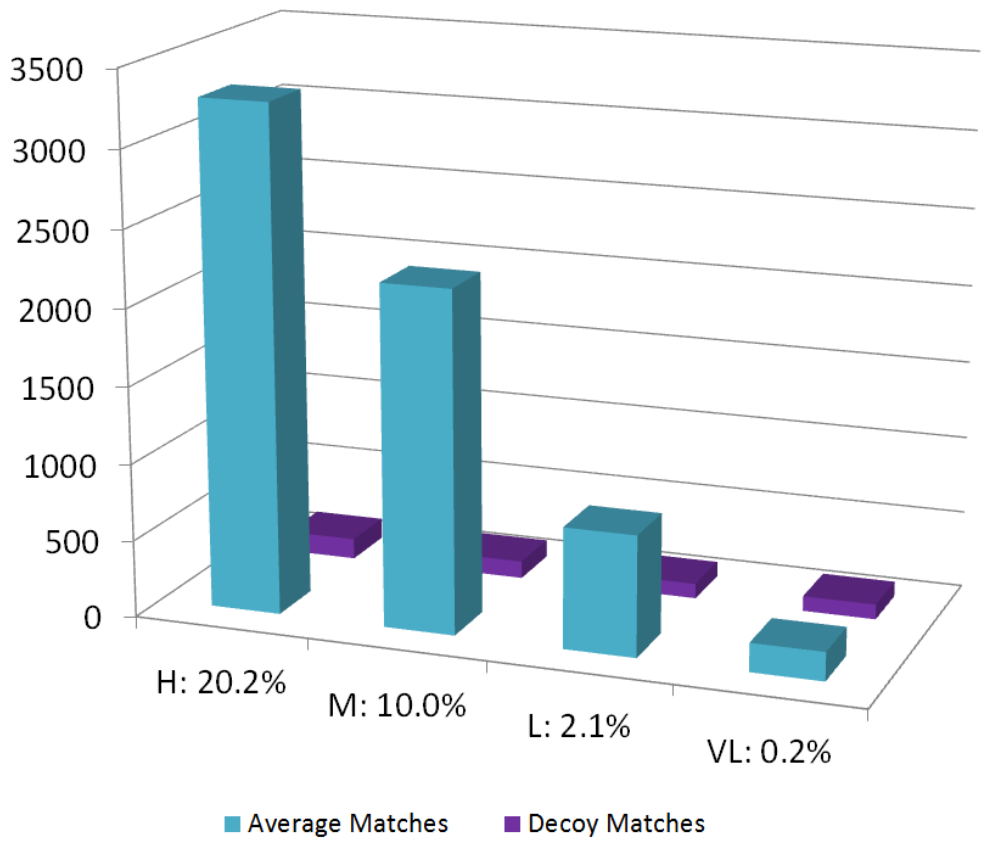


Figure V.3. Average matches for high (H), medium (M), low (L), and very low (VL) target (*Spiroplasma kunkelii*) percentage mock sample databases (MSDs) queried with *S. kunkelii*-specific e-probes and decoy e-probes using the EDNA pipeline and parsed at 10^{-3}.

Table V.2. Query of lowest target percentage mock sample database (0.01%; VL-2) with *Spiroplasma kunkelii*-specific e-probes and decoy e-probes using the EDNA pipeline. BLASTn outputs were parsed at E-values of 10^{-9}, 10^{-6}, and 10^{-3}.

Parser E-value	Matches	Hits	Decoy Matches	Decoy Hits	Chi-square	P-value
e^{-9}	0	0	0	0	-	-
e^{-6}	1	1	0	0	1.000	0.31731
e^{-3}	99	136	102	126	0.046	0.83018

Leafhopper Acquisition and RNA Extraction

E. exitiosus surviving the 25-day latent period were tested for acquisition of *S. kunkelii* by PCR. Two prior test replications demonstrated that no negative control leafhoppers acquired *S. kunkelii* (data not shown); negative control tests were therefore reduced. One representative of four to five leafhoppers from each control plant was tested to ensure spiroplasma absence from the *S. kunkelii*-naïve control groups. Each remaining leafhopper in the *S. kunkelii*-infected test group was assayed for presence of *S. kunkelii*. All four control group leafhoppers were free of spiroplasma infection (Figure V.4). Twelve of 37 test group *E. exitiosus* were clearly infected with *S. kunkelii* (Figure V.4). An additional 7 individuals from the test group showed very faint amplification of *S. kunkelii*. These individuals were not included for RNA extraction, however, because infection status was low or uncertain. Ten individuals (pool of 5 males and 5 females) were randomly chosen from both *S. kunkelii*-naïve and *S. kunkelii*-positive (Figure V.4: Males - 7A, 11A, 16A, 5B, and 13B; Females - 9A, 10A, 3B, 4B, and 15B) groups for RNA extraction. Approximately 9 µg RNA (within purity ranges required by Cofactor Genomics for sequencing) per sample was submitted for sequencing.

Transcriptome Sequencing and EDNA Detection

Illumina NextSeq500 single-end sequencing of *S. kunkelii*-infected and *S. kunkelii*-naïve total RNA generated 35,717,463 and 32,760,587 1x75 (single-end 75-base) reads, respectively. Results of the datasets query with *S. kunkelii* e-probes and decoy e-probes parsed at 10^{-9}, 10^{-6}, and 10^{-3} are shown in Table V.3. Detection of *S. kunkelii* in the infected dataset was significant ($p < 0.00001$) at each parser E-value. *S. kunkelii* was not detected in the naïve dataset at any of the parser E-values (Table V.3). At 10^{-3}, *S. kunkelii* e-probes had 1 match and 1 hit in the naïve dataset; however, the decoy e-probes showed 5 matches and 5 hits. Statistically, this detection in the naïve dataset at this E-value was insignificant ($p = 0.10226$).

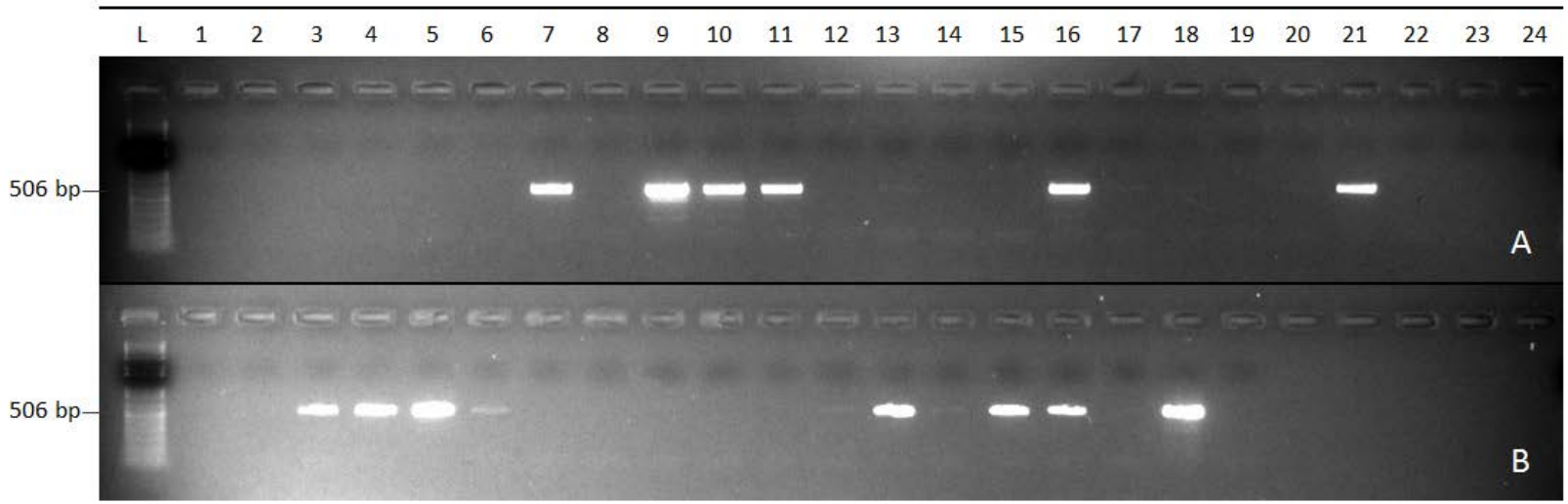


Figure V.4. Amplification of a 506-bp segment of the *Spiroplasma kunkelii* spiralin gene in individual *Exitianus exitiosus*. Amplicons were separated by electrophoresis in 1.5% agarose gels in 1X TAE. Lanes 1-4A contain *S. kunkelii*-naïve control group individuals; lanes 5-24A and 1-17B contain individuals from the *S. kunkelii*-infected test group; lanes 18B and 19B contain positive (*S. kunkelii* culture) and negative (nanopure water) controls respectively.

Table V.3. EDNA detection of *Spiroplasma kunkelii* in raw RNA transcriptome sequence NGS datasets derived from infected and naïve *Exitianus exitiosus*. Detection results are parsed at 10^{-9}, 10^{-6}, and 10^{-3} for each dataset.

Dataset	E-value	E-probe Matches	E-probe Hits	Decoy Matches	Decoy Hits	Chi-square	P-value
Infected	e^{-9}	530	1035	0	0	558.10	< 0.00001
Naïve		0	0	0	0	--	--
Infected	e^{-6}	617	1264	0	0	655.41	< 0.00001
Naïve		0	0	0	0	--	--
Infected	e^{-3}	1400	4786	5	5	1598.38	< 0.00001
Naïve		1	1	5	5	2.67	0.10226

Traditional Bioinformatic Analysis

Four assemblies were created for the two transcriptome datasets (Table V.4). In the *S. kunkelii*-infected dataset, 363,056 reads (1.02%) mapped to the *S. kunkelii* chromosomal and plasmid genomes using Read Mapper, and in the *S. kunkelii*-naïve dataset, 245,964 reads (0.75%) mapped to the chromosomal genome only. Using Bowtie 2, 183,902 reads (0.51%) in the *S. kunkelii*-infected dataset mapped to the chromosomal and plasmid genomes, while 10,462 reads (0.03%) mapped to the chromosomal genome only in the *S. kunkelii*-naïve dataset. Considering only these data, it would be assumed that either *S. kunkelii* sequences were present within the *S. kunkelii*-naïve dataset or that these were false positives; therefore, these results were further analyzed using BLAST and MEGAN5.

BLAST files resulting from reads mapping to the *S. kunkelii* genomes were analyzed with MEGAN5. Metagenomic analysis showed that sequences assigned within class Mollicutes were present in abundance only in the *S. kunkelii*-infected *E. exitiosus* datasets (Figures V.5 and V.6). *S. kunkelii*-naïve datasets resulted in zero sequences within the Bowtie 2 assembled reads (0.0%; Table V.5) and six sequences in the Read Mapper assembled reads (0.002%; Table V.5) assigned within Mollicutes. Of the six sequences, one was assigned to genus *Mycoplasma*, two to *Mycoplasma canis*, one to *Mycoplasma* sp. 31848, one to family Spiroplasmataceae, and one to *Spiroplasma*; no sequences were assigned to *S. kunkelii*. The two assembly programs showed relatively similar numbers of reads assigned to each taxonomic rank (starting with phylum Tenericutes, class Mollicutes, order Entomoplasmatales, family Spiroplasmataceae, genus *Spiroplasma*, species *S. kunkelii*; Table V.6), indicating some consistency between the programs to map sequenced *S. kunkelii* reads within the insect datasets to the *S. kunkelii* genomes.

Table V.4. Assemblies of *Spiroplasma kunkelii*-infected and *S. kunkelii*-naïve *Exitianus exitiosus* transcriptomes. Reads were mapped to the *S. kunkelii* chromosomal genome and plasmid genomes using Geneious Read Mapper and Bowtie 2 Assemblers.

Assembler	*E. exitiosus* Dataset	*S. kunkelii* Genome	# Assembled Reads	% Pairwise Identity
Read Mapper	Infected	Chromosomal	362,254	90.2
		pSKU226	530	97.9
		pSKU205	253	97.9
		pSKU76	19	99.6
	Naïve	Chromosomal	245,964	93.9
Bowtie 2	Infected	Chromosomal	183,152	96.9
		pSKU226	499	98.7
		pSKU205	242	98.5
		pSKU76	9	99.6
	Naïve	Chromosomal	10,461	98.7

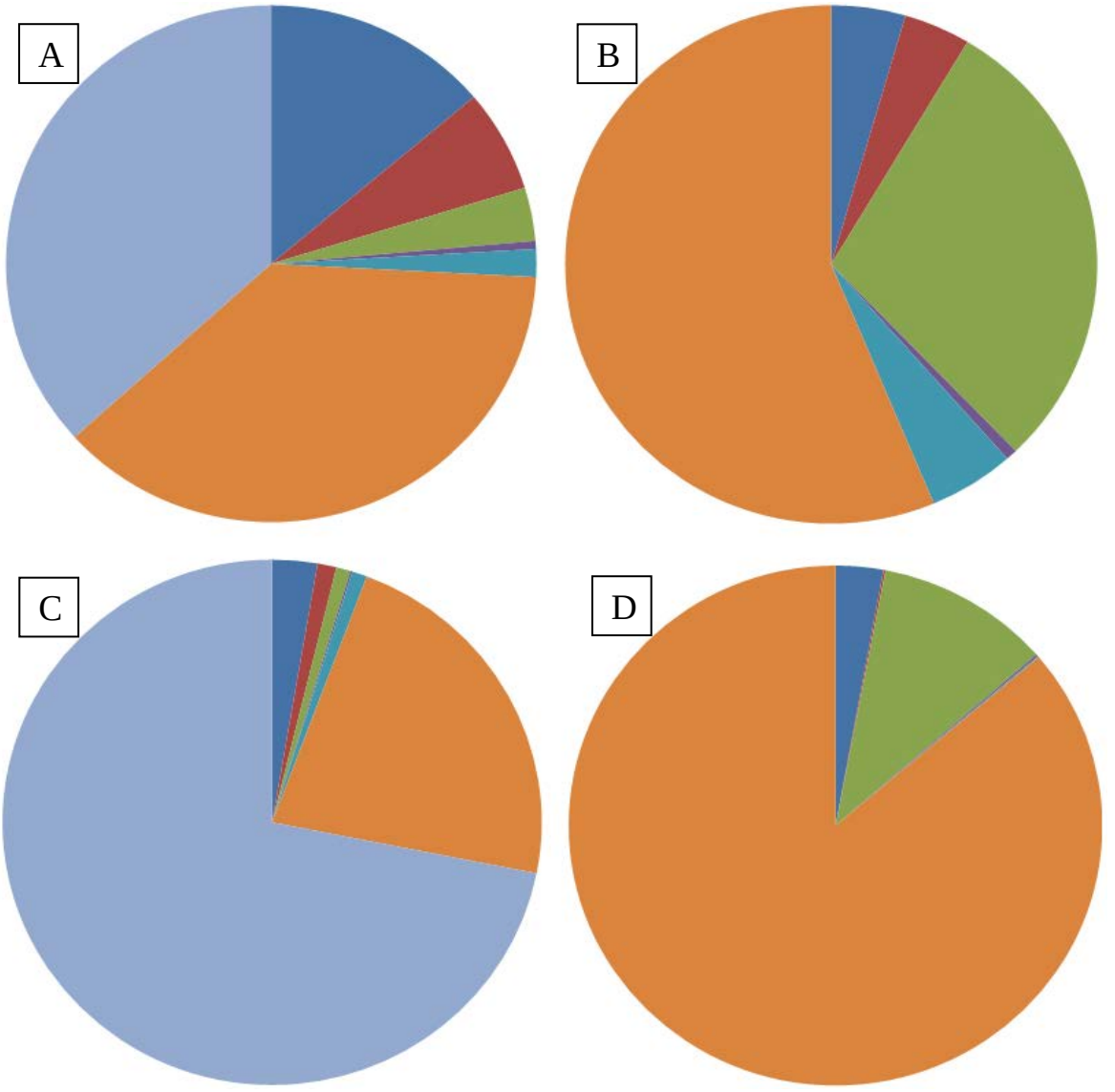


Figure V.5. MEGAN5 analysis of reads assembled from *Spiroplasma kunkelii*-infected and *Spiroplasma kunkelii*-naïve *Exitianus exitiosus* transcriptome sequence datasets to the *S. kunkelii* chromosomal and plasmid genomes by Read Mapper and Bowtie 2. Results shown are the proportions of reads assigned within specific Classes (Mollicutes, Gammaproteobacteria, Betaproteobacteria, Flavobacteriia, or others), higher order taxonomic rankings, or unassigned. (A) Read Mapper assembled *Spiroplasma kunkelii*-infected reads; (B) Read Mapper assembled *Spiroplasma kunkelii*-naïve reads; (C) Bowtie 2 assembled *Spiroplasma kunkelii*-infected reads; (D) Bowtie 2 assembled assembled *Spiroplasma kunkelii*-naïve reads.

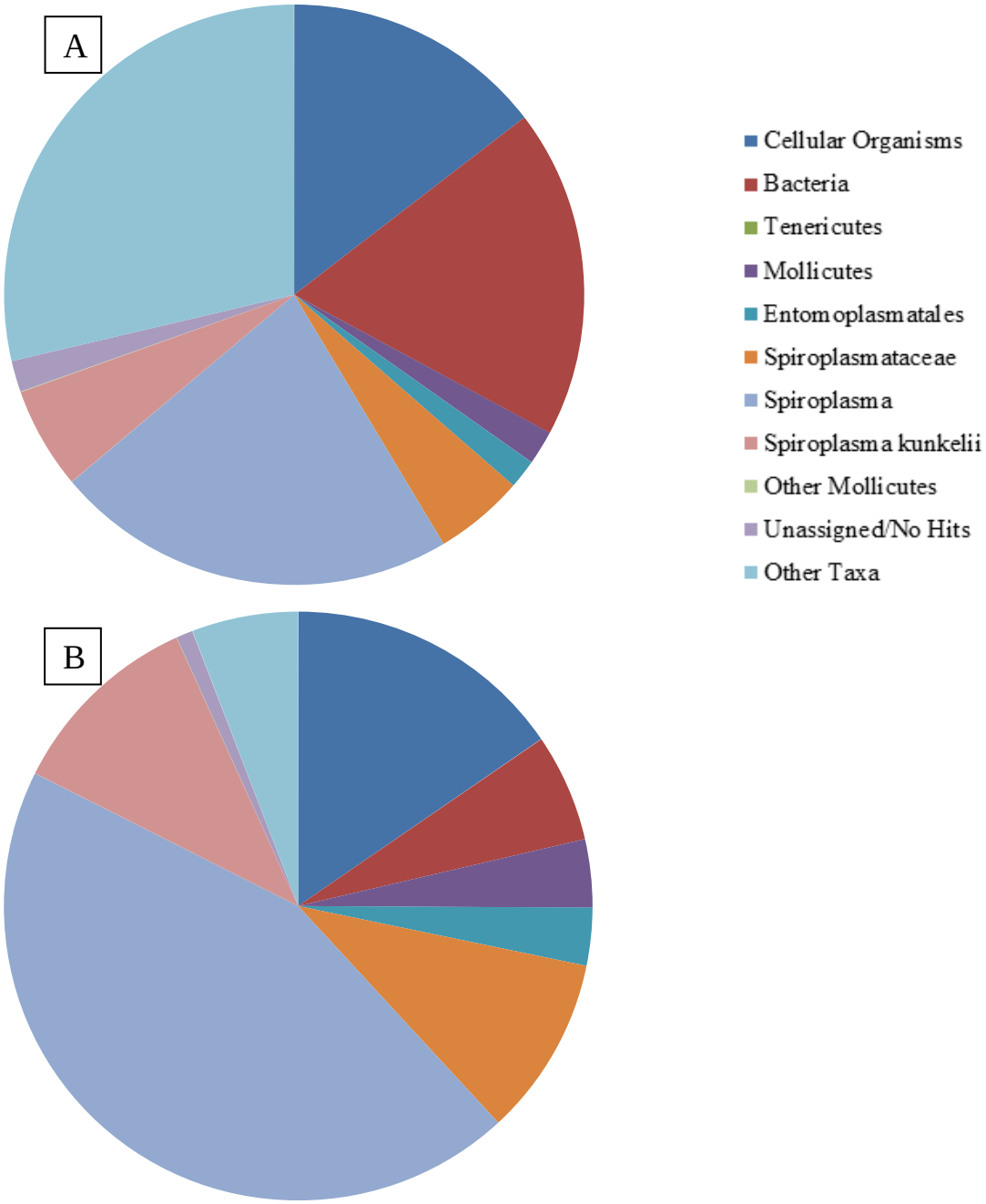


Figure V.6. MEGAN5 analysis of reads assembled from *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcriptome sequence datasets to the *S. kunkelii* chromosomal and plasmid genomes by Read Mapper and Bowtie 2. Results shown are the proportions of reads assigned to specific taxonomic rankings. (A) Read Mapper assembled *Spiroplasma kunkelii*-infected reads; (B) Bowtie 2 assembled *Spiroplasma kunkelii*-infected reads.

Table V.5. Percentages of *S. kunkelii*-infected and *S. kunkelii*-naïve *E. exitiosus* transcriptome sequencing reads assembled to *S. kunkelii* genomes by Read Mapper and Bowtie 2 that were assigned within Mollicutes versus other classes or higher order taxonomic groups.

	Read Mapper		**Bowtie 2**	
	***Sk*-infected**	***Sk*-naïve**	***Sk*-infected**	***Sk*-naïve**
Mollicutes	36.7%	0.002%	72.0%	0.0%
Gammaproteobacteria	13.8	4.5	2.7	2.9
Betaproeobacteria	6.4	4.1	1.1	0.1
Flavobacteriia	3.3	29.2	0.9	10.6
Other Classes	0.5	0.7	0.1	0.1
Unassigned/No Hits	1.7	5.2	0.9	0.1
Cellular/Kingdom/Phylum	37.5	56.3	22.3	86.2

Table V.6. Taxonomic profiles of the *S. kunkelii*-infected *E. exitiosus* transcriptome reads assembled to *S. kunkelii* genomes by Read Mapper and Bowtie 2 with focus on the taxonomic lineage of *S. kunkelii*. Taxonomic groups in bold indicate *S. kunkelii*-specific lineage groups that are absent or assigned in extremely low abundance in the assembled *S. kunkelii*-naïve *E. exitiosus* datasets.

Taxonomic Group	**Read Mapper**		**Bowtie 2**	
	Read Count	**Percentage (%)**	**Read Count**	**Percentage (%)**
Cellular Organisms	52822	14.55	28253	15.36
Bacteria	66536	18.33	10856	5.90
Tenericutes	1	0.00028	1	0.00054
Mollicutes	6901	1.90	6831	3.71
Entomoplasmatales	5868	1.62	5821	3.17
Spiroplasmataceae	18134	4.99	18078	9.83
Spiroplasma	81623	22.48	81185	44.15
Spiroplasma kunkelii	20609	5.68	20411	11.10
Other Mollicutes	121	0.033	49	0.027
Unassigned/No Hits	6289	1.73	1696	0.92
Other Taxa	104152	28.69	10721	5.83
Total	363056	100	183902	100

Disscussion

The import and introduction of exotic insects poses significant risks to agriculture, horticulture, and forestry. An average of 35,000 insects are intercepted at ports of entry every year (McCullough et al., 2006), and this number is on the rise. Nearly 40% of intersected insects are Hemipterans, which include many phytopathogen vectors. For many vectors, the risk posed to plants lies primarily in their capacity to transmit plant pathogens rather than in damage caused by feeding. The routine identification of intercepted insects on plant materials by morphological characteristics is essential to preventing the introduction of invasive vectors; however, high throughput screening of these insects and the pathogens they might be carrying could facilitate enhanced understanding of the movement of particular insects and pathogens and, in turn, provide for more robust and timely introduction interventions.

NGS significantly advanced the field of genomics by theoretically enabling the sequencing of any and potentially all organisms within a given sample (Mardis, 2008). This capability provides a substantial opportunity to diagnostics, including the plant pathogen detection and disease diagnosis. To adapt this technology to rapid processing needs at ports of entry, the EDNA bioinformatics tool was developed (Stobbe et al., 2013). Initially developed, tested, and validated for the detection of plant pathogens within or on plants, this study provides the first validation of EDNA-based pathogen detection within an animal metatranscriptome. Specifically, *S. kunkelii* was detected significantly within the transcriptome of *S. kunkelii*-infected *E. exitiosus* using designed *S. kunkelii*-specific e-probes and the EDNA pipeline without any prior traditional processing of sequencing reads. After designing target-specific e-probes, detection of a pre-determined target within raw genomic or transcriptomic NGS datasets can be made within minutes using EDNA.

In previous EDNA studies, total nucleic acids were amplified (whole genome and whole transcriptome amplification) prior to sequencing in attempt to increase target sequences within the dataset (Stobbe et al., 2014; Blagden et al., 2016), but the effects of this strategy on pathogen sequence abundance within tested datasets were not described. This step was omitted in this study in order to avoid introducing a bias for or against pathogen sequences and to allow for an accurate representation of pathogen sequence abundance, assuming no sequencing bias, within the datasets. Only transcriptomic sequencing, as opposed to both transcriptomic RNA and genomic DNA sequencing, was performed in this study. While previous studies showed sensitive detection using EDNA in genomic datasets, DNA detection does not necessarily indicate a live or active pathogen infection. Detection in transcriptomes, however, provides greater evidence for a functioning pathogen.

MSD generation and testing provide estimates of e-probe detection sensitivity. Comparing the numbers of *S. kunkelii* e-probe matches resulting from *S. kunkelii*-infected dataset query (530, 671, and 1400 matches for e^{-9}, e^{-6}, and e^{-3}, respectively) to those from MSD query, a low to medium level, between 3.4-6.7% (calculated from MSD match trendlines; Appendix 5), of target sequences was initially estimated within the *S. kunkelii*-infected dataset. However, considering the number of sequences that mapped to the *S. kunkelii* chromosomal and plasmid genomes, the target percentage in the *S. kunkelii*-infected dataset is actually much lower. The two infected dataset assemblies resulted in 363,056 (Read Mapper) and 183,902 (Bowtie 2) total *S. kunkelii* genomes-aligned reads. Of these assembled reads, 133,256 and 132,376, respectively, (calculated from Table V.6; summed read counts for Tenericutes through Other Mollicutes) were determined to be of Mollicutes origin, indicating that only approximately 0.37% (both 133,256 Read Mapper assembled reads and 132,376 Bowtie 2 assembled reads divided by 35,717,463 total *S. kunkelii*-infected reads) of the dataset's total reads were likely from *S. kunkelii*. This disparity in MSD estimates and traditional analysis results indicates that e-probes could be more

sensitive in practice than in theory or that actual results are not directly comparable to MSD results. In this study, MSDs were generated to simulate 454 sequencing runs; however, leafhopper samples were sequenced using Illumina. The differences between 454 and Illumina sequence results may explain this inconsistency.

Two assembly programs were used to map and reference-align the *S. kunkelii*-infected and *S. kunkelii*-naïve *E. exitiosus* reads to the *S. kunkelii* genomes. This strategy accounted for some variation in results that would be observed among different read mapping programs. Geneious Read Mapper provided a user-friendly graphical interface based mapping program, while Bowtie 2 is a commonly used program that can be performed both via the Geneious software or linux-based computational analysis. Different settings were used with the two programs to account for lower (Read Mapper) to higher (Bowtie 2) levels of alignment stringency, which resulted in nearly double the number of reads shown to align to the reference genome by Read Mapper than Bowtie 2. A rapid method of *S. kunkelii* detection in the datasets could include merely running a mapping program; however, without further BLAST and metagenomic analysis of these results, thousands of reads in both the infected and naïve datasets could have been assumed to be of *S. kunkelii* origin merely because they mapped to the *S. kunkelii* reference genomes. According to these results, if all reads assigned by BLAST and MEGAN5 within the Tenericutes taxonomic branch are assumed to derive from the *S. kunkelii* infection, based on zero (Bowtie 2) and only six (Read Mapper) total reads assigned to this branch within the *S. kunkelii*-naïve dataset, then only 72.0% (132,376 reads determined to be of Mollicutes origin divided by 183,902 total Bowtie 2 assembled reads) or 36.7% (133,256 reads of Mollicutes origin divided by 363,056 total Read Mapper assembled reads), respectively, of reference-mapped reads actually resulted from *S. kunkelii* sequencing. In contrast to traditional bioinformatic analysis methods, EDNA provides a rapid, target-focused, accurate method of pathogen detection within raw NGS datasets.

EDNA is a valuable tool for pathogen detection within next generation sequencing datasets and is applicable to diagnostic, discovery, and general research purposes. Validated for pathogen detection within plants and insects, the pipeline can be applied to identification of any target pathogen within various organismal backgrounds, which could be particularly useful for screening samples collected at ports of entry.

REFERENCES

Alivizatos, A. S. 1982. Feeding behavior of the spiroplasma vectors *Dlbulus maidis* and *Euscelidius variegatus in vivo* and *in vitro*. Chronika Benaki Phytopathological Institutu 13:128-44.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Barros, T. S., Davis, R. E., Resende, R. O., and Dally, E. L. 2001. Design of a polymerase chain reaction for specific detection of corn stunt spiroplasma. Plant Disease, 85(5):475-80.

Blagden, T., Schneider, W., Melcher, U., and Fletcher, J. 2013. *In silico* Adaptation of EDNA (E-probe Diagnostic Nucleic Acid Analysis) for Detection of Foodborne Pathogens. Annual Meeting, July 28-31, 2013, IAFP.

Blagden, T., Schneider, W.L., Melcher, U., Daniels, J., Fletcher, J. 2016. Adaptation and validation of E-probe diagnostic nucleic acid analysis for detection of *Escherichia coli* O157:H7 in metagenomic data of complex food matrices. Journal of Food Protection 79:574-581

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674-3676.

Daniels, J., Schneider, B., and Ochoa-Corona, F. Detection of prokaryotic plant pathogens using next generation sequencing and electronic probes (EDNA). In preparation for submission to BMC Microbiology.

El Hadidi, M., Ruscheweyh, H. J. and Huson, D. 2013. Improved metagenome analysis using MEGAN5. In Joint 21st Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) and 12th European Conference on Computational Biology (ECCB).

Espindola, A., Schneider, W., Hoyt, P. R., Marek, S. M., and Garzon, C. 2015. A new approach for detecting fungal and oomycete plant pathogens in next generation sequencing metagenome data utilising electronic probes. International journal of data mining and bioinformatics 12(2):115-128.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and

Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647-1649.

Langmead, B. and Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.

Maramorosch, K. and Harris, K. ed., 1979. Leafhopper vectors and plant disease agents. Elsevier.

Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. Trends in genetics 24(3):133-141.

McCullough, D. G., Work, T. T., Cavey, J. F., Liebhold, A. M., and Marshall, D. 2006. Interceptions of nonindigenous plant pests at US ports of entry and border crossings over a 17-year period. Biological invasions 8(4):611-630.

Nault, L. R. 1980. Maize bushy stunt and corn stunt: a comparison of disease symptoms, pathogen host ranges, and vectors. Phytopathology 70:659-662.

Richter, D.C., Ott, F., Auch, A.F., Schmid, R., and Huson, D.H. 2008. MetaSim—A Sequencing Simulator for Genomics and Metagenomics. PLoS ONE 3(10):e3373. doi:10.1371/journal.pone.0003373

Satya, R. V., Zavaljevski, N., Kumar, K., and Reifman, J. 2008. A high-throughput pipeline for designing microarray-based pathogen diagnostic assays. BMC Bioinformatics 9(1):1.

Stobbe, A.H, Daniels, J., Espindola, A.S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J., and Schneider, W. 2013. E-probe diagnostic nucleic acid analysis (EDNA): a theoretical approach for handling of next generation sequencing data for diagnostics. Journal of Microbiological Methods 94:356-366.

Stobbe, A., W. Schneider, and U. Melcher. 2013. A bioinformatic pipeline for use in metagenomic virus discovery. Annual Meeting. American Phytopathological Society. Abstract.

Stobbe, A.H., Schneider, W.L., Hoyt, P.R. and Melcher, U. 2014. Screening Metagenomic Data for Viruses Using the E-Probe Diagnostic Nucleic Acid Assay. Phytopathology 104(10):1125-1129.

CHAPTER VI

DETECTION OF AN IMMUNE RESPONSE IN GRAY LAWN LEAFHOPPER, *EXITIANUS EXITIOSUS*, TRANSCRIPTOMES AFTER ACQUISITION OF *SPIROPLASMA KUNKELII*

Abstract

Many questions still remain about the interactions between phytopathogenic spiroplasmas and their insect vectors. Both reduced and increased fitness have been reported in plant-infecting spiroplasma-insect vector systems, indicating that spiroplasmas have intimate molecular interactions with their vectors. To begin understanding the vector's response to acquisition of spiroplasmas at a transcriptomic level, the transcriptomes of *Spiroplasma kunkelii*-infected and naïve gray lawn leafhoppers, *Exitianus exitiosus* (Uhl.), were compared. Total RNA was extracted from three replicate pools of male and female leafhoppers 25 days (latent period) after a 24-hour acquisition access period from *S. kunkelii*-inoculated feeding sachets. Three similar replicates were simultaneously completed for the naïve leafhopper control groups after exposure to *S. kunkelii*-negative feeding sachets. Poly-A selected RNA was sequenced using the Illumina NextSeq500 platform, and *de novo* assembly was performed by the Trinity pipeline. After dataset normalization, transcript abundances were estimated and differentially expressed transcripts were identified and annotated. Differential expression of over 300 transcripts, including immune defense related genes, was observed in *E. exitiosus* infected with *S. kunkelii*.

Introduction

Insects in the order Hemiptera threaten global agricultural production by being phytophagous and having specialized mouthparts that readily allow for direct inoculation of pathogens into plant vasculature and other tissues. Aphids (Aphididae), whiteflies (Aleyrodidae), leafhoppers (Cicadellidae), planthoppers (Fulgoromorpha), and other hemipterans account for most plant pathogen vectors, collectively transmitting hundreds of viruses, many phytopathogenic bacteria, and all known plant-infecting spiroplasmas and phytoplasmas (Hogenhout et al., 2008; Weintraub and Beanland, 2006). Mobile vectors play important roles in the epidemiology of many of these pathogens, serving as intermediate hosts between plants and accounting for much of their long-distance movement (Ossiannilsson, 1966; Purcell, 1982). Dissecting the relationships among plant hosts, pathogens, and insect vectors is crucial to disease management.

Interactions between Cicadellids and spiroplasmas are of particular importance because leafhoppers comprise the majority of phytopathogenic spiroplasma vectors and can serve as the primary vehicles of pathogen spread throughout an agricultural field (Fletcher et al., 1998). There are only three known plant-infecting spiroplasmas, *Spiroplasma citri*, *S. kunkelii*, and *S. phoeniceum*; however, several leafhopper species are demonstrated vectors, including natural and experimental vectors (Nault, 1980; O'Hayer et al., 1983; Carloni et al., 2011). Though transmission by experimental vector species has been demonstrated in the laboratory rather than observed in nature, the potential for natural transmission certainly exists; these species merely have not yet been recorded as the culpable vector species in any disease outbreaks.

Some interesting behavioral differences have been identified between the natural and experimental vectors of phytopathogenic spiroplasma species; namely, plant-infecting

spiroplasma infections in natural vectors have been shown to have either neutral or even beneficial effects on vector fitness, whereas infections in experimental vectors have demonstrated negative effects on vector fitness (Madden and Nault, 1983; Madden et al., 1984; Ebbert and Nault, 1994). The many non-phytopathogenic spiroplasma species are known to have a diversity of relationships with their arthropod hosts, including commensal, mutualistic, and pathogenic (Regassa and Gasparich, 2006). This diversity appears to be the case for plant pathogenic spiroplasmas with different vector species as well, and the variety of relationships could be related to long co-evolution, or lack thereof, of spiroplasma species with their insect vector hosts (Maramorosch, 1981; Madden and Nault. 1983; Bandi et al., 2001).

There is conflicting evidence for whether or not spiroplasmas induce an immune response in insects. *S. poulsonii,* a *Drosophila* associated species that is vertically transmitted from mother to progeny and kills males during embryogenesis, and a closely related male-killing spiroplasma did not induce any common antimicrobial responses (Toll and Imd pathways) after infection in Drosophila (Hurst et al., 2003; Anbutsu and Fukatsu, 2010; Herren and Lemaitre, 2011). Injection of plant pathogenic *S. citri* into *Drosophila melanogaster* flies did not induce an innate immune response either (Herren and Lemaitre, 2011). However, immune-related gene activation has recently been dectected in *Circulifer haematoceps,* a primary vector of *S. citri* in Europe, in response to *S. citri* infection (Eliautout et al., 2016).

Knowledge of the interactions between mollicutes and their insect hosts, particularly the molecular effects of phytopathogenic spiroplasma infections on their leafhopper vectors, is severely lacking. Understanding these complex relationships at a genetic level could lead to the development of more vector transmission control options. Comparing the transcriptomes of spiroplasma-infected and spiroplasma-naïve vectors, specifically for differentially expressed transcripts, can reveal genes that are important to spiroplasma infection and, potentially,

transmission. To this end, the transcriptomes of *S. kunkelii*-infected and *S. kunkelii*-naïve gray lawn leafhoppers, *Exitianus exitiosus* (Uhl.) (Hemiptera: Cicadellidae), were assembled and evaluated for differential expression in response to *S. kunkelii* presence. This vector system was chosen for representation of response in an experimental vector, which could be compared to that of a primary vector (Eliautout et al., 2016) and evaluated for responses indicative of the negative effects of spiroplasmas on non-primary vectors (Madden and Nault, 1983; Madden et al., 1984; Ebbert and Nault, 1994).

Materials and Methods

S. kunkelii Acquisition by *E. exitiosus*

Third-fourth instar *E. exitiosus* were collected from bermudagrass in late September, 2014 in Stillwater, OK. Leafhoppers were transferred in groups of 10 to feeding sachets with 500 µL of D10 medium with 1% fetal bovine serum (FBS) (negative control; *S. kunkelii*-naive) or 500 µL of D10 medium with resuspended *S. kunkelii* (*S. kunkelii*-infected) (Alivizatos, 1982). Feeding sachets consisted of a 1 oz. medicine cup covered with two stretched parafilm membranes with D10 medium or D10/*S. kunkelii* between the layers. *S. kunkelii* CR2-3X cultures were grown to log phase at 29˚C in LD8A3 medium (Maramorosch and Harris, 1979) and evaluated by a dark-field microscope for quality. Cultures were centrifuged at 10,000 rpm for 1.5 min to form a pellet and medium supernatant was discarded. *S. kunkelii* were then resuspended in D10 medium to approximately 10^8 CFU/mL for feeding sachet acquisition. After a 24 hr acquisition access period (AAP), leafhoppers were transferred to healthy corn (*Zea mays* L.) at the 2-4 leaf stage for a latent period (LP) of 25 days to allow enough time for *S. kunkelii* to be acquired and begin to multiply in the vector (Nault, 1980). Corn and leafhoppers were maintained

in a growth chamber with a 16 hr photoperiod at 27°C (day) and 18°C (night). Three total replications were conducted to produce three *S. kunkelii*-naïve and three *S. kunkelii*-infected groups for sequencing and analysis.

Surviving *E. exitiosus* leafhoppers were tested by PCR for *S. kunkelii* acquisition and to ensure *S. kunkelii* absence in the control group. Leafhoppers were collected from corn, briefly chilled, and the right metathoracic leg was excised. Legs were placed individually in 180 µL sterilized phosphate buffered saline (1x PBS; pH 7.4) and homogenized with a plastic pestle, and DNA was extracted using the Qiagen DNeasy® Blood and Tissue Kit (Valencia, CA) following the manufacturer's instructions. *S. kunkelii* was then detected by endpoint PCR using *S. kunkelii*-specific primers F1 and R1, which amplify a segment of the spiralin gene (Barros et al., 2001). PCR reactions consisted of 10 µL GoTaq® Green Master Mix (Promega, Madison, WI), 1 µL 5 mM F1, 1 µL 5 mM R1, 2 µL template DNA (Qiagen extraction; total DNA concentration not measured), and 6 µL nanopure water. Positive control reactions consisted of *S. kunkelii* template DNA extracted from cultures, and negative controls consisted of nanopure water (no template control; NTC). Cycling conditions used were 95°C for 3 min; 35 cycles of 95°C for 30 sec, 60°C for 1 min, 72°C for 45 sec; and 72°C for 5 min. Amplicons were visualized by gel electrophoresis in a 1.5 % agarose gel in 1XTAE with 3 µL SYBR® Safe DNA Gel Stain (Invitrogen, Carlsbad, CA) per 100 mL gel.

RNA Extraction and Sequencing

In replicates 1 and 2, total RNA was extracted from a pool of five male and five female *E. exitiosus* for both the *S. kunkelii*-naïve and *S. kunkelii*-infected groups using Trizol and Purelink™ RNA Mini Kit (Invitrogen) per the manufacturer's instructions. For replicate 3, RNA was extracted from two males and two females (discussed in results) using the same protocol. Resulting RNA was cleaned with deoxyribonuclease I (amplification grade; Invitrogen) treatment

to remove contaminating DNA and quantified three times using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE) before being shipped on dry ice to Cofactor Genomics (4044 Clayton Ave. Saint Louis, MO) for sequencing.

RNA was processed for library construction by Cofactor Genomics. Briefly, poly-adenylated RNA was selected by oligo-dT beads and sheared to appropriate size for cDNA synthesis. Double-stranded cDNA was end-repaired and A-tailed for adaptor ligation. Indexed adaptors were ligated to sample cDNA, and the adaptor-ligated cDNA was size-selected on a 2% SizeSelect™ E-Gel (Invitrogen) and amplified by PCR. Library size and quality was assessed on an Agilent Bioanalyzer and library yield was quantified by qPCR using the Kapa Biosystems library quantification kit (Wilmington, MA) prior to sequencing on the Illumina NextSeq500 (San Diego, CA) as 1x75 base reads following manufacturer's protocols.

Transcriptome Assembly and Differential Expression Analysis

A *de novo E. exitiosus* transcriptome assembly was generated by first combining all six read datasets. Initially combining all datasets for assembly allows for generation of the most complete transcriptome assembly possible, providing a backbone that represents the complexity of the transcriptome (e.g. representing isoforms that may be present in only one dataset or in only the control or test group). The Trinity (Grabherr et al., 2011) version 2.1.1 assembly pipeline was performed on the combined reads dataset. The Trinity pipeline consists of three software phases. Briefly, in the first phase, Jellyfish (Marçais and Kingsford, 2011) is used to count k-mers from input reads and build a k-mer catalog. The Trinity Inchworm module then contructs linear contigs, assembling unique transcripts from the k-mers. In phase two, Chrysalis, contigs are clustered, each cluster representing the transcriptional complexity of a gene (alternative splicing and unique sequences of paralogous genes), and de Bruijn graphs are contructed for each cluster. In phase three, Butterfly, de Bruijn graphs are analyzed, and all full-length transcripts for

alternative isoforms and paralogous genes are reported. All three Trinity phases were performed to compile an assembled transcriptome.

After assembly, RSEM (RNA-Seq by Expectation Maximization) was performed to estimate transcript abundances (Li and Dewey, 2011). RSEM, which automates the use of Bowtie (Langmead et al., 2009) for mapping and alignment of reads back to the assembled trancriptome and normalizes read counts, was executed on all six datasets individually as described by the manual. Resulting transcript abundance estimates for each dataset were then used for differential expression analysis using edgeR (Robinson et al., 2010), which reports the logFC (log_2 fold change) and logCPM (log_2 count per million) for each transcript in the exposed group compared against the control group as well as the statistical significance of the results, including the p-value and false discovery rate (FDR).

Annotation

Protein coding regions within the transcripts were predicted by running TransDecoder (http://transdecoder.github.io/), which identified candidate open reading frames (ORFs) at a minimum 100 amino acids in length. BLAST analysis was performed on candidate proteins using blastp and blastx algorithms for query against the NCBI non-redundant (nr) protein database. The top differentially expressed transcripts identified by edgeR (FDR < 0.05; p-value < 0.00012) were additionally analyzed using the Blast2GO PRO platform (version 3.3; Conesa et al., 2005). Transcripts were queried against the NCBI nr database using blastx at an E-value threshold of $1x10^{-3}$. Protein functional analysis was performed by running InterProScan (EMBL-EBI; Jones et al., 2014) against all supported databases to identify common protein signatures including protein families, domains, repeats, and sites. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) analyses were also performed on differentially expressed transcripts using Blast2GO.

Results

S. kunkelii Acquisition by *E. exitiosus*

After the 25-day latent period, surviving *E. exitiosus* were tested by PCR for acquisition of *S. kunkelii*. Subsets of at least one in five of the *S. kunkelii*-naïve leafhoppers were tested to confirm the absence of *S. kunkelii* in control groups. All tested controls were free of *S. kunkelii* infection (Figures VI.1-VI.3). All leafhoppers remaining in the *S. kunkelii*-infected test group were assayed for *S. kunkelii* presence. Of 37 leafhoppers tested in *S. kunkelii*-infected replicate 1 group, 12 produced strong bands for *S. kunkelii* spiralin amplification, indicating positive infection (Figure VI.1). Seven additional individuals showed faint amplification. In replicate 2, 12 of 25 were clearly positive for *S. kunkelii*, and lesser amplification was again observed for eight individuals (Figure VI.2). In replicate 3, six of 20 had strong amplification, and only one had weak amplification (Figure VI.3).

Individuals with weak amplification were not selected for RNA extraction because infection status was low or uncertain. For replicates 1 and 2, ten individuals (pool of five males and five females) were chosen randomly from both *S. kunkelii*-naïve and *S. kunkelii*-positive groups for RNA extraction (Figure VI.1: Males - 7A, 11A, 16A, 5B, and 13B; Females - 9A, 10A, 3B, 4B, and 15B and Figure VI.2: Males – 7A, 1B, 6B, 7B, and 10B; Females – 1A, 5A, 6A, 11A, and 8B). Because only four males and two females positively acquired for replicate 3, four individuals were pooled for RNA extraction (two males and two females; Figure VI.3: Males – 8A and 5B; Females – 9A and 11B). Approximately 9 µg of RNA (within purity ranges required by Cofactor Genomics for sequencing) was submitted per sample for sequencing.

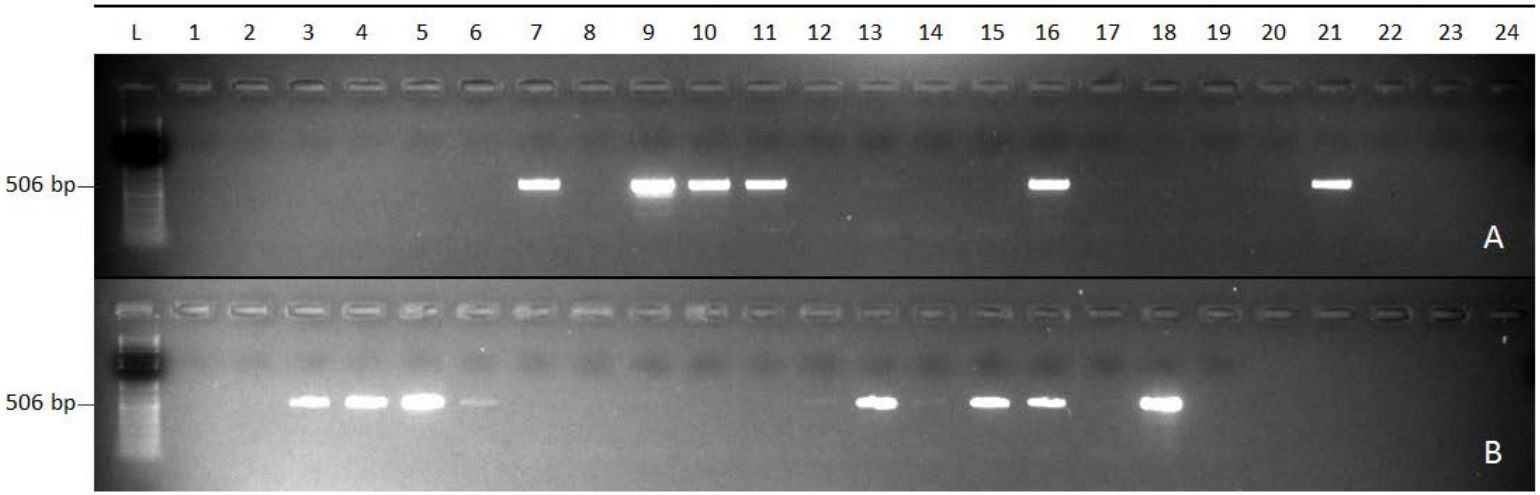


Figure VI.1. Amplification of a portion of the *Spiroplasma kunkelii* spiralin gene in *S. kunkelii*-infected and *S. kunkelii*-naïve Replicate 1 group *Exitianus exitiosus*. Lanes 1-5A represent negative control *S. kunkelii*-naïve individuals, lanes 6-24A and 1-17B represent *S. kunkelii*-infected test group individuals, lane 18B contains the positive control (culture *S. kunkelii*), and lane 19B contains the negative control (nanopure water; NTC).

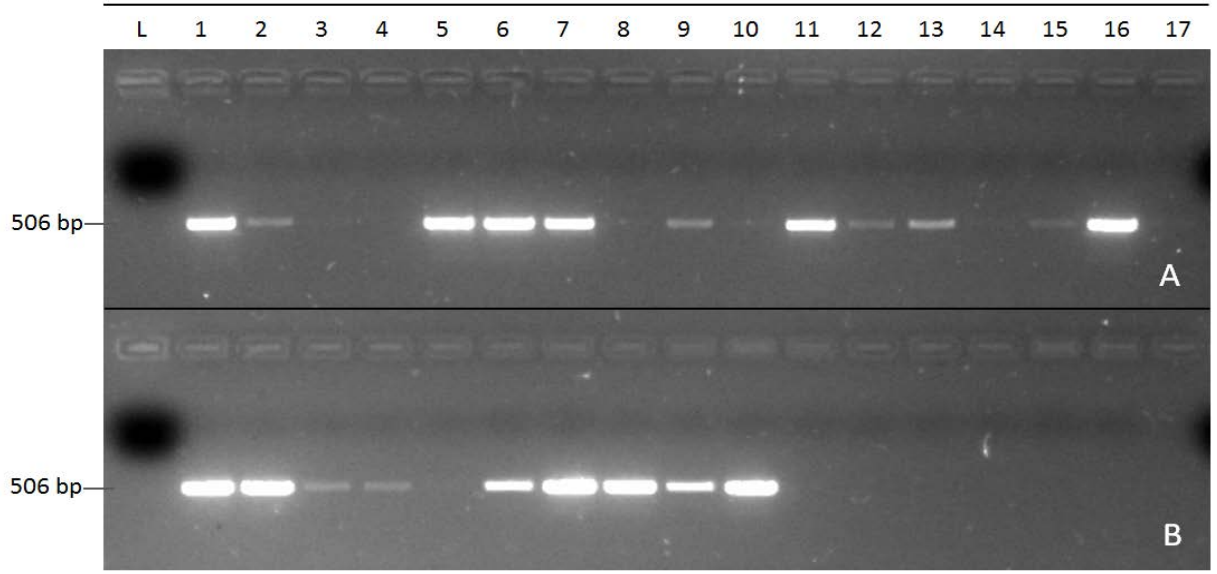


Figure VI.2. Amplification of a portion of the *Spiroplasma kunkelii* spiralin gene in *S. kunkelii*-infected and *S. kunkelii*-naïve Replicate 2 group *Exitianus exitiosus*. Lanes 1-15A and 1-10B represent *S. kunkelii*-infected test group individuals, lanes 11-16B represent negative control *S. kunkelii*-naïve individuals, lane 16A contains the positive control (culture *S. kunkelii*), and lane 17B contains the negative control (nanopure water; NTC).

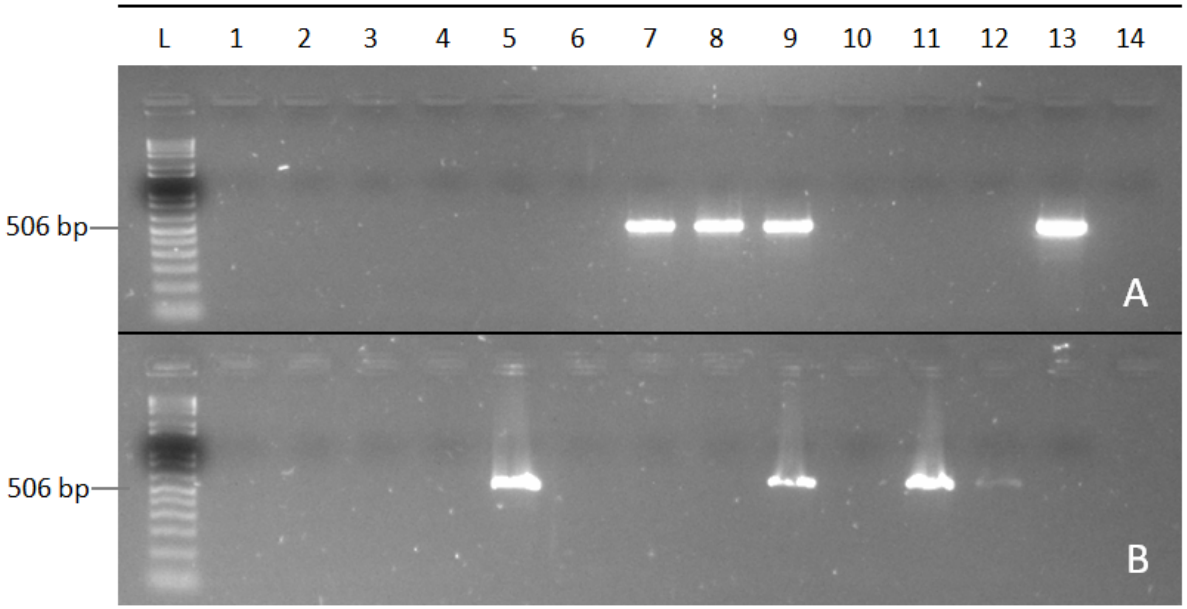


Figure VI.3. Amplification of a portion of the *Spiroplasma kunkelii* spiralin gene in *S. kunkelii*-infected and *S. kunkelii*-naïve Replicate 3 group *Exitianus exitiosus*. Lanes 1-5A represent negative control *S. kunkelii*-naïve individuals, lanes 6-12A and 1-13B represent *S. kunkelii*-infected test group individuals, lane 13A contains the positive control (culture *S. kunkelii*), and lane 14A contains the negative control (nanopure water; NTC).

Sequencing and Differential Expression

Illumina NextSeq500 single-end sequencing of three replicates of *S. kunkelii*-infected and *S. kunkelii*-naïve mRNA-selected RNA samples generated six sequence datasets (Table VI.1). Assembly of the transcriptome from combined datasets resulted in the generation of 97,712 predicted genes and 136,042 transcripts. GC content was 39.10%. Considering all generated transcript contigs, the average contig length was 614.74 bases, median contig length was 345 bases, contig N50 was 931 bases, and the total number of assembled bases was 83,629,995. TransDecoder predicted 52,521 open reading frames translating at least 100 amino acid long polypeptides.

Normalized transcript counts (expected counts), TPM (transcripts per million), and FPKM (fragments per kilobase million) values were reported for each transcript in each dataset using RSEM. These results for the three *S. kunkelii*-infected datasets were grouped and analyzed against grouped *S. kunkelii*-naïve datasets for differential expression using edgeR. Differential transcript expression in the *S. kunkelii*-infected group as compared to *S. kunkelii*-naïve was demonstrated at varying levels of fold change and statistical support. MA and volcano plots were generated for dataset expression representation (Figures VI.4 and VI.5). A total of 308 transcripts were highly supported as differentially expressed with an FDR (false discovery rate) less than 5% (p-value < 0.00012). Transcripts upregulated from the *S. kunkelii*-naïve control totaled 173, and downregulated transcripts totaled 135. LogFC ($\log_2$ fold change) ranged from 2.27 to 11.64 for upregulated transcripts and -2.11 to -13.61 for downregulated transcripts.

Table VI.1. Summary statistics of *Sprioplasma kunkelii*-infected and *S. kunkelii*-naïve polyA-selected RNA Illumina NextSeq500 single-end 75-base sequencing.

Sample	Read Count	# Bases Generated
Sk-naïve 1	28,596,543	2,144,740,725
Sk-naïve 2	27,215,643	2,041,173,225
Sk-naïve 3	27,389,988	2,054,249,100
Sk-infected 1	27,514,995	2,063,624,625
Sk-infected 2	26,952,809	2,021,460,675
Sk-infected 3	28,413,436	2,131,007,700

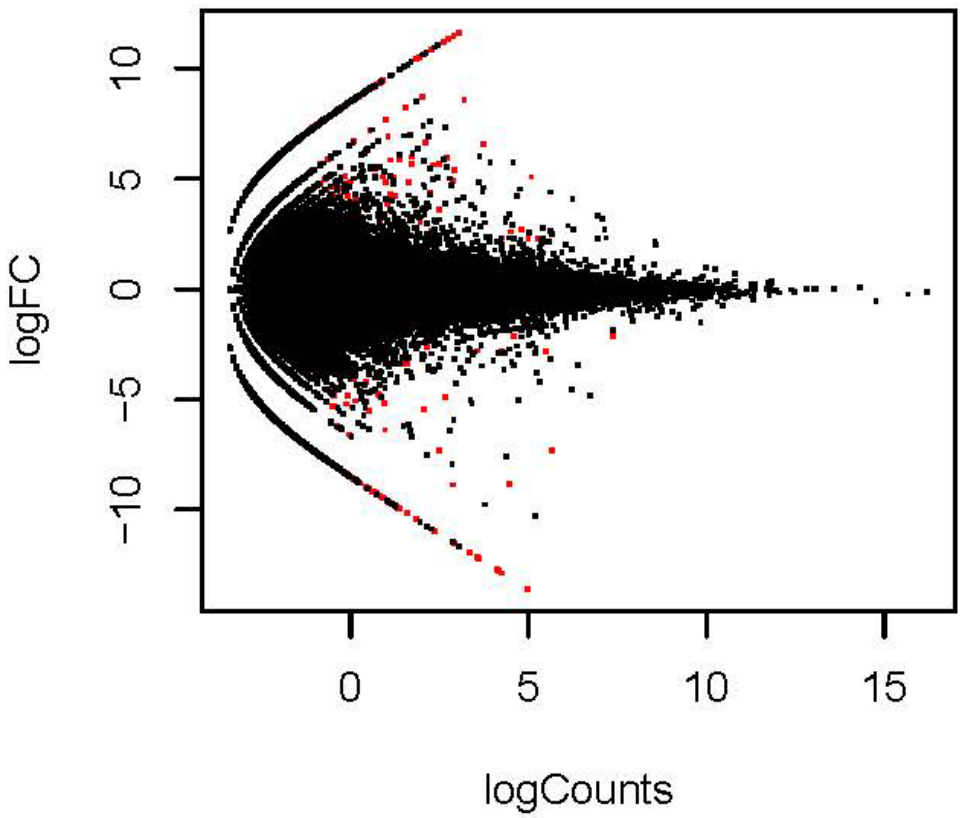


Figure VI.4. 'MA' plot of log ratio (FC: fold change) versus abundance (transcript counts) of *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcript expression. Each dot represents a transcript, and differentially expressed transcripts (False Discovery Rate < 0.05) are plotted in red.

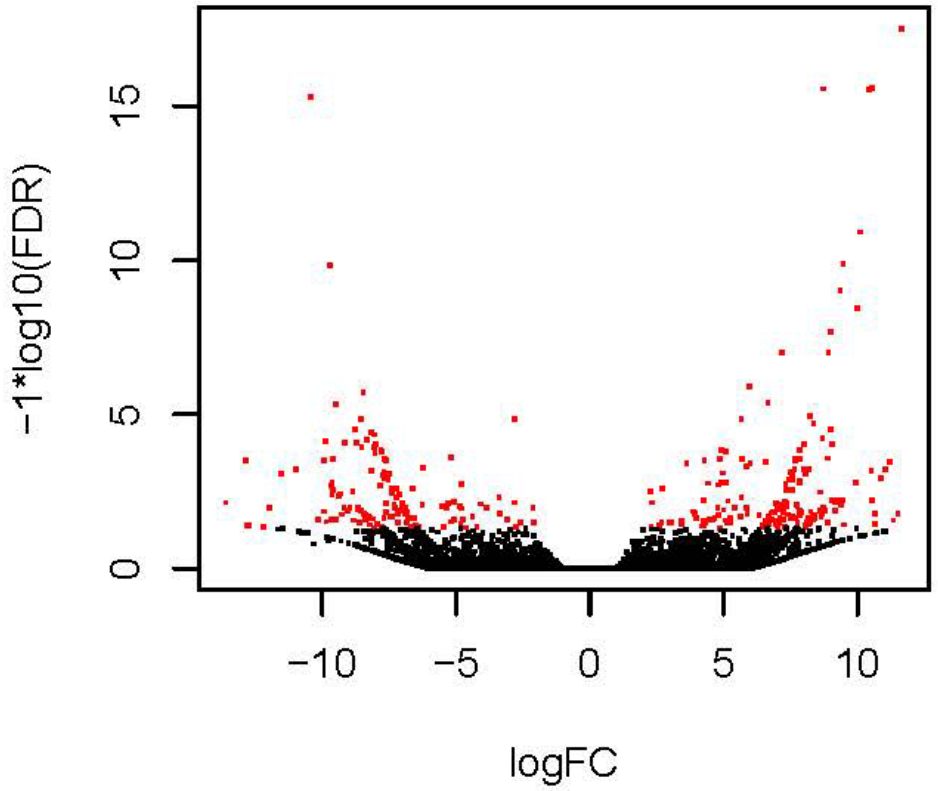


Figure VI.5. Volcano plot of log significance versus fold change of *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcript expression. Each dot represents a transcript, and differentially expressed transcripts (False Discovery Rate < 0.05) are plotted in red.

Differential Expression Annotation

Significant blastx results (E-value < $1x10^{-3}$; nr database) were returned for 72 of 173 (41.6%) upregulated transcripts and 60 of 135 (44.47%) downregulated transcripts. Top hit results are summarized in Table VI.2 (upregulated) and Table VI.3 (downregulated) for sequences producing more than a 'hypothetical protein' description. Sequences producing no blastx hits were then queried with blastn against the nucleotide (nt) database at an E-value threshold of $1x10^{-3}$. These queries produced 10 (upregulated) and 12 (downregulated) additional hits. Protein functional signatures (IPS) were identified for 73 of 173 (42.2%) upregulated transcripts and 68 of 135 (50.4%) transcripts. GO annotations were returned for 24 of 173 (13.9%) upregulated and 23 of 135 (17.0%) downregulated transcripts. Protein families were identified in transcripts producing IPS results (Tables VI.4). GO terms identified for differentially regulated transcripts were classified into three domains including cellular component, molecular function, and biological process (Tables VI.5 – VI.9).

Eight of the upregulated transcripts had Rfam homology to bacterial rRNA. Blastn analysis of these 8 transcripts showed these sequences to be of *S. kunkelii* origin. No *S. kunkelii*-predicted sequences were observed in the downregulated group. To identify all *S. kunkelii*-related sequences, all transcripts were then mapped to the *S. kunkelii* chromosomal (Reference Sequence NZ_CP010899.1) and plasmid genomes (pSKU226: NZ_CP012423.1; pSKU205: NZ_CP012424.1; and pSKU76: NZ_CP012425.1) available in Genbank using Bowtie. Eighteen of the assembled transcripts mapped to the *S. kunkelii* chromosomal genome. All eighteen trancripts aligned with regions of the 16S, 23S, or 5S rRNA genes (gb|DQ319068.1).

Transcripts with homology to viruses were differentially expressed. Transcripts with high homology to *Nilaparvata lugens* (brown planthopper) honeydew virus 1 (E-value: $1.65E^{-19}$) (Accession BAN19725) and *Sendai virus* (E-value: 0) (Accession BAM62828) were significantly

Table VI.2. Blastx top hit descriptions for upregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts.

Sequence Name	Length	Blastx Top Hit Description (HSP)	Accession	E-Value
TRINITY_DN41553_c0_g4_i1	208	gi|501290646|dbj|BAN20225.1|small heat shock protein hsp20 family [Riptortus pedestris]	BAN20225	1.03E-35
TRINITY_DN44680_c0_g1_i2	2763	gi|951535075|ref|XP_014471099.1|PREDICTED: phosphofurin acidic cluster sorting protein 2 isoform X3 [Dinoponera quadriceps]	XP_014471099	0
TRINITY_DN66689_c0_g1_i1	330	gi|899696901|gb|KMY34488.1|hydrolase, partial [Lysinibacillus xylanilyticus]	KMY34488	4.45E-24
TRINITY_DN12071_c0_g1_i1	384	gi|758344886|dbj|GAN11851.1|hydrolase, partial [Mucor ambiguus]	GAN11851	5.01E-39
TRINITY_DN3682_c0_g1_i1	235	gi|758345115|dbj|GAN11783.1|beta-lactamase [Mucor ambiguus]	GAN11783	1.06E-25
TRINITY_DN43832_c2_g1_i2	891	gi|730371165|gb|KHJ42742.1|Beta-ketoacyl synthase protein [Trichuris suis]	KHJ42742	2.80E-17
TRINITY_DN39908_c0_g5_i1	231	gi|775460946|dbj|BAQ94506.1|NcSP26 [Nephotettix cincticeps]	BAQ94506	4.92E-07
TRINITY_DN51854_c0_g1_i2	3048	gi|805820280|ref|XP_012150225.1|PREDICTED: scm-like with four MBT domains protein 1 isoform X2 [Megachile rotundata]	XP_012150225	0
TRINITY_DN48862_c0_g1_i10	802	gi|671758636|gb|AII97728.1|BLTX341 [Nephila pilipes]	AII97728	3.01E-04
TRINITY_DN57333_c0_g1_i1	374	gi|922983695|ref|WP_053391285.1|chromosome partitioning protein [Spiroplasma kunkelii]	WP_053391285	8.09E-58
TRINITY_DN56499_c0_g1_i1	12305	gi|939650050|ref|XP_014274301.1|PREDICTED: hemocytin [Halyomorpha halys]	XP_014274301	2.13E-05
TRINITY_DN46488_c0_g1_i2	1875	gi|751225531|ref|XP_011166154.1|PREDICTED: abl interactor 2 isoform X1 [Solenopsis invicta]	XP_011166154	2.06E-163
TRINITY_DN51716_c0_g1_i1	2635	gi|913321724|ref|XP_013192038.1|PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like [Amyelois transitella]	XP_013192038	1.70E-114
TRINITY_DN35783_c0_g1_i2	2612	gi|954560521|ref|XP_014604035.1|PREDICTED: axonemal 84 kDa protein-like [Polistes canadensis]	XP_014604035	5.62E-142
TRINITY_DN49793_c0_g1_i4	3503	gi|641654099|ref|XP_008179070.1|PREDICTED: piggyBac transposable element-derived protein 4-like [Acyrthosiphon pisum]	XP_008179070	2.09E-123
TRINITY_DN55915_c0_g1_i2	1559	gi|358336229|dbj|GAA54788.1|retrovirus-related Pol polyprotein from transposon 17.6, partial [Clonorchis sinensis]	GAA54788	3.09E-06
TRINITY_DN19653_c0_g1_i1	730	gi|913324481|ref|XP_013193561.1|PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like isoform X1 [Amyelois transitella]	XP_013193561	4.70E-44
TRINITY_DN52762_c3_g2_i2	934	gi|939642190|ref|XP_014271579.1|PREDICTED: collagenase 3-like [Halyomorpha halys]	XP_014271579	1.94E-64
TRINITY_DN48784_c0_g1_i7	1948	gi|939279207|ref|XP_014260174.1|PREDICTED: UPF0160 protein MYG1, mitochondrial [Cimex lectularius]	XP_014260174	7.71E-121
TRINITY_DN51841_c0_g2_i1	600	gi|939253065|ref|XP_014246192.1|PREDICTED: calcium-independent protein kinase C-like [Cimex lectularius]	XP_014246192	5.90E-64
TRINITY_DN54921_c0_g1_i1	246	gi|939650851|ref|XP_014274569.1|PREDICTED: pancreatic lipase-related protein 2-like [Halyomorpha halys]	XP_014274569	1.07E-38
TRINITY_DN31834_c0_g8_i1	581	gi|646713044|gb|KDR17545.1|40S ribosomal protein S28 [Zootermopsis nevadensis]	KDR17545	3.97E-26
TRINITY_DN53277_c0_g1_i3	1861	gi|646721793|gb|KDR23023.1|ADP-dependent glucokinase [Zootermopsis nevadensis]	KDR23023	0
TRINITY_DN55986_c0_g2_i1	207	gi|936695411|ref|XP_014225687.1|PREDICTED: eukaryotic peptide chain release factor GTP-binding subunit ERF3A isoform X3 [Trichogramma pretiosum]	XP_014225687	5.15E-38
TRINITY_DN56120_c1_g1_i1	1490	gi|646721126|gb|KDR22596.1|hypothetical protein L798_12726, partial [Zootermopsis nevadensis]	KDR22596	2.23E-58
TRINITY_DN52938_c1_g1_i2	3804	gi|939659146|ref|XP_014277639.1|PREDICTED: protein zer-1 homolog [Halyomorpha halys]	XP_014277639	0

Table VI.2 (cont.). Blastx top hit descriptions for upregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts.

Sequence Name	Length	Blastx Top Hit Description (HSP)	Accession	E-Value
TRINITY_DN44879_c0_g1_i1	995	gi|646711868|gb|KDR16898.1|TCDD-inducible poly [ADP-ribose] polymerase [Zootermopsis nevadensis]	KDR16898	1.47E-32
TRINITY_DN44927_c0_g1_i11	1564	gi|646707070|gb|KDR13990.1|Peroxisomal membrane protein PEX16 [Zootermopsis nevadensis]	KDR13990	4.78E-100
TRINITY_DN69978_c0_g2_i1	538	gi|939277071|ref|XP_014259035.1|PREDICTED: myosin light chain alkali isoform X2 [Cimex lectularius]	XP_014259035	1.52E-21
TRINITY_DN28360_c0_g1_i1	323	gi|496526990|dbj|BAN19725.1|polyprotein [Nilaparvata lugens honeydew virus 1]	BAN19725	1.65E-19
TRINITY_DN49040_c0_g1_i1	1691	gi|939651145|ref|XP_014274687.1|PREDICTED: protein regulator of cytokinesis 1-like isoform X1 [Halyomorpha halys]	XP_014274687	4.63E-20
TRINITY_DN43917_c0_g1_i2	1988	gi|646699403|gb|KDR10033.1|Chromobox protein-like protein 3 [Zootermopsis nevadensis]	KDR10033	2.42E-24
TRINITY_DN34014_c0_g1_i2	360	gi|759068441|ref|XP_011343723.1|PREDICTED: longitudinals lacking protein, isoforms A/B/D/L-like isoform X4 [Cerapachys biroi]	XP_011343723	2.05E-13
TRINITY_DN44905_c0_g1_i1	277	gi|766919065|ref|XP_011501947.1|PREDICTED: defensin-like [Ceratosolen solmsi marchali]	XP_011501947	5.90E-05
TRINITY_DN56048_c3_g1_i2	374	gi|847136778|ref|XP_012820565.1|PREDICTED: zinc finger MYM-type protein 1-like [Xenopus (Silurana) tropicalis]	XP_012820565	4.48E-09
TRINITY_DN52715_c0_g1_i8	1330	gi|93279411|pdb|2F2L|XChain X, Crystal Structure Of Tracheal Cytotoxin (Tct) Bound To The Ectodomain Complex Of Peptidoglycan Recognition Proteins Lca (Pgrp-Lca) And Lcx (Pgrp-Lcx)	2F2L_X	2.12E-07
TRINITY_DN53185_c0_g1_i2	686	gi|944461796|gb|KQL98670.1|hypothetical protein Y1Q_030100 [Alligator mississippiensis]	KQL98670	4.80E-26
TRINITY_DN48378_c0_g3_i2	221	gi|322792874|gb|EFZ16707.1|hypothetical protein SINV_11620, partial [Solenopsis invicta]	EFZ16707	3.15E-43
TRINITY_DN52943_c1_g2_i4	1558	gi|238800254|gb|ACR56003.1|Ly-6/neurotoxin superfamily member 1 [Nilaparvata lugens]	ACR56003	1.48E-35
TRINITY_DN51877_c0_g1_i2	329	gi|954571118|ref|XP_014609752.1|PREDICTED: ubiquinone biosynthesis protein COQ9, mitochondrial [Polistes canadensis]	XP_014609752	9.84E-47
TRINITY_DN46801_c0_g1_i1	3415	gi|646714069|gb|KDR18162.1|Dual 3',5'-cyclic-AMP and -GMP phosphodiesterase 11 [Zootermopsis nevadensis]	KDR18162	5.90E-71
TRINITY_DN27375_c0_g2_i1	806	gi|482875755|gb|AGK40905.1|i-type lysozyme 2 [Nilaparvata lugens]	AGK40905	6.91E-51
TRINITY_DN17529_c0_g1_i1	686	gi|749754956|ref|XP_011139936.1|PREDICTED: longitudinals lacking protein, isoforms A/B/D/L isoform X48 [Harpegnathos saltator]	XP_011139936	1.34E-11
TRINITY_DN38434_c0_g1_i1	3674	gi|408684321|dbj|BAM62828.1|phosphoprotein [Sendai virus]	BAM62828	0

Table VI.3. Blastx top hit descriptions for downregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts.

Sequence Name	Length	Blastx Top Hit Description (HSP)	Accession	E-Value
TRINITY_DN49350_c0_g1_i3	3300	gi|907678727|ref|XP_013106010.1|PREDICTED: uncharacterized protein LOC106086048 [Stomoxys calcitrans]	XP_013106010	1.91E-10
TRINITY_DN51787_c0_g1_i4	1588	gi|906466664|gb|KNC28062.1|hypothetical protein FF38_04439 [Lucilia cuprina]	KNC28062	2.69E-101
TRINITY_DN49350_c0_g1_i1	2263	gi|907678727|ref|XP_013106010.1|PREDICTED: uncharacterized protein LOC106086048 [Stomoxys calcitrans]	XP_013106010	1.07E-10
TRINITY_DN38336_c0_g4_i1	272	gi|91090342|ref|XP_967194.1|PREDICTED: holotricin-1 [Tribolium castaneum]	XP_967194	2.71E-05
TRINITY_DN51246_c1_g1_i3	1466	gi|970918938|ref|XP_015126136.1|PREDICTED: T-complex protein 11-like protein 1 isoform X2 [Diachasma alloeum]	XP_015126136	3.80E-82
TRINITY_DN53294_c0_g1_i4	2084	gi|915660791|gb|KOC63503.1|Protein cornichon like protein 4 [Habropoda laboriosa]	KOC63503	1.21E-62
TRINITY_DN27835_c0_g1_i1	204	gi|776195612|ref|YP_009129265.1|polyprotein [Graminella nigrifrons virus 1]	YP_009129265	5.93E-24
TRINITY_DN54253_c0_g1_i4	3006	gi|646717385|gb|KDR20260.1|Peroxisomal Lon protease-like protein 2 [Zootermopsis nevadensis]	KDR20260	0
TRINITY_DN21136_c0_g1_i1	674	gi|70909531|emb|CAJ17188.1|ribosomal protein S13e [Georissus sp. APV-2005]	CAJ17188	4.12E-100
TRINITY_DN56143_c1_g1_i3	5728	gi|939262121|ref|XP_014250975.1|PREDICTED: dynamin isoform X11 [Cimex lectularius]	XP_014250975	0
TRINITY_DN52250_c1_g1_i2	487	gi|156554739|ref|XP_001605381.1|PREDICTED: carbonic anhydrase 2-like [Nasonia vitripennis]	XP_001605381	2.56E-16
TRINITY_DN51660_c0_g1_i8	3708	gi|641669163|ref|XP_008184709.1|PREDICTED: E3 ubiquitin-protein ligase ZNRF2 [Acyrthosiphon pisum]	XP_008184709	3.26E-87
TRINITY_DN54130_c2_g2_i3	307	gi|328723284|ref|XP_001948577.2|PREDICTED: spermine oxidase-like isoform X2 [Acyrthosiphon pisum]	XP_001948577	2.95E-17
TRINITY_DN1918_c0_g1_i1	206	gi|776195612|ref|YP_009129265.1|polyprotein [Graminella nigrifrons virus 1]	YP_009129265	1.03E-19
TRINITY_DN55760_c0_g1_i1	2576	gi|328708796|ref|XP_001952676.2|PREDICTED: uncharacterized protein LOC100160086 isoform X1 [Acyrthosiphon pisum]	XP_001952676	1.13E-10
TRINITY_DN54794_c0_g1_i1	838	gi|919031261|ref|XP_013399249.1|PREDICTED: UNC93-like protein MFSD11 isoform X2 [Lingula anatina]	XP_013399249	6.89E-21
TRINITY_DN40623_c0_g2_i1	265	gi|776195612|ref|YP_009129265.1|polyprotein [Graminella nigrifrons virus 1]	YP_009129265	9.50E-24
TRINITY_DN69978_c0_g1_i1	550	gi|939277071|ref|XP_014259035.1|PREDICTED: myosin light chain alkali isoform X2 [Cimex lectularius]	XP_014259035	6.51E-17
TRINITY_DN55302_c2_g2_i11	1778	gi|826458684|ref|XP_012533758.1|PREDICTED: RNA-binding protein 45 [Monomorium pharaonis]	XP_012533758	1.16E-99
TRINITY_DN43832_c2_g1_i9	563	gi|328703183|ref|XP_001944653.2|PREDICTED: fatty acid synthase-like [Acyrthosiphon pisum]	XP_001944653	3.18E-15
TRINITY_DN15956_c0_g1_i1	361	gi|776195612|ref|YP_009129265.1|polyprotein [Graminella nigrifrons virus 1]	YP_009129265	2.18E-45
TRINITY_DN18220_c0_g2_i1	654	gi|749754956|ref|XP_011139936.1|PREDICTED: longitudinals lacking protein, isoforms A/B/D/L isoform X48 [Harpegnathos saltator]	XP_011139936	2.26E-14
TRINITY_DN19444_c0_g1_i1	256	gi|776195612|ref|YP_009129265.1|polyprotein [Graminella nigrifrons virus 1]	YP_009129265	4.32E-35
TRINITY_DN55306_c0_g1_i1	6318	gi|861620675|gb|KMQ87338.1|reverse ribonuclease integrase, partial [Lasius niger]	KMQ87338	9.24E-123
TRINITY_DN26978_c0_g1_i2	503	gi|512915301|ref|XP_004928356.1|PREDICTED: speckle targeted PIP5K1A-regulated poly(A) polymerase-like [Bombyx mori]	XP_004928356	1.10E-09
TRINITY_DN54845_c3_g1_i10	1501	gi|646691680|gb|KDR07187.1|Beta-sarcoglycan [Zootermopsis nevadensis]	KDR07187	7.95E-113

Table VI.3 (cont.). Blastx top hit descriptions for downregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts.

Sequence Name	Length	Blastx Top Hit Description (HSP)	Accession	E-Value
TRINITY_DN53861_c0_g2_i2	1830	gi|456358985|dbj|BAM93347.1|doublesex [Trypoxylus dichotomus]	BAM93347	1.20E-24
TRINITY_DN33379_c0_g1_i2	469	gi|776195612|ref|YP_009129265.1|polyprotein [Graminella nigrifrons virus 1]	YP_009129265	1.37E-06
TRINITY_DN52088_c0_g1_i5	823	gi|646722394|gb|KDR23407.1|Carbohydrate sulfotransferase 5, partial [Zootermopsis nevadensis]	KDR23407	4.55E-54
TRINITY_DN46900_c1_g1_i2	720	gi|328703972|ref|XP_001942950.2|PREDICTED: fatty acid synthase-like isoform X2 [Acyrthosiphon pisum]	XP_001942950	2.58E-21
TRINITY_DN55672_c2_g2_i6	1389	gi|817202742|ref|XP_012277164.1|PREDICTED: phosphatidylinositol-binding clathrin assembly protein LAP isoform X3 [Orussus abietinus]	XP_012277164	1.80E-09
TRINITY_DN32374_c0_g2_i1	250	gi|776195612|ref|YP_009129265.1|polyprotein [Graminella nigrifrons virus 1]	YP_009129265	4.75E-14
TRINITY_DN52120_c0_g2_i3	1563	gi|646717300|gb|KDR20209.1|putative protein phosphatase 2C T23F11.1 [Zootermopsis nevadensis]	KDR20209	2.01E-176
TRINITY_DN44461_c0_g1_i8	1114	gi|939691298|ref|XP_014289660.1|PREDICTED: ankyrin-3-like isoform X2 [Halyomorpha halys]	XP_014289660	1.54E-170
TRINITY_DN43560_c0_g2_i6	1424	gi|646714741|gb|KDR18595.1|NAD-dependent deacetylase sirtuin-3, mitochondrial [Zootermopsis nevadensis]	KDR18595	3.31E-79
TRINITY_DN19653_c0_g2_i1	727	gi|913324481|ref|XP_013193561.1|PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like isoform X1 [Amyelois transitella]	XP_013193561	3.36E-46
TRINITY_DN26245_c0_g2_i1	565	gi|723941620|gb|AIY24631.1|peptidoglycan recognition protein SD, partial [Graminella nigrifrons]	AIY24631	5.29E-30
TRINITY_DN55278_c1_g1_i1	963	gi|641667664|ref|XP_008184171.1|PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like [Acyrthosiphon pisum]	XP_008184171	1.09E-17
TRINITY_DN55700_c0_g2_i9	2620	gi|952513954|gb|KRT80450.1|membrane transporter [Oryctes borbonicus]	KRT80450	6.13E-154
TRINITY_DN33856_c0_g1_i2	225	gi|768407098|ref|XP_011599713.1|PREDICTED: LOW QUALITY PROTEIN: 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-3-like [Aquila chrysaetos canadensis]	XP_011599713	3.79E-04
TRINITY_DN31695_c0_g1_i1	304	gi|646708194|gb|KDR14596.1|hypothetical protein L798_11176 [Zootermopsis nevadensis]	KDR14596	1.29E-16
TRINITY_DN50081_c0_g1_i1	415	gi|755957921|ref|XP_011303913.1|PREDICTED: MD-2-related lipid-recognition protein-like [Fopius arisanus]	XP_011303913	3.70E-20

Table VI.4. Homologous protein families identified by InterProScan using Blast2GO for significantly upregulated and downregulated (FDR < 0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts. EMBL-EBI families are provided in parentheses (https://www.ebi.ac.uk/interpro/).

Upregulated Transcripts	Downregulated Transcripts
Peroxisome membrane protein, Pex16 (IPR013919)	Major facilitator superfamily (IPR011701)
Phosphofurin acidic cluster sorting protein 1 (IPR019381)	Lon protease (IPR027065)
Microtubule-associated protein (IPR007145)	DMRT/protein doublesex/protein male abnormal 3 (IPR026607)
ADP-specific phosphofructokinase/glucokinase (IPR007145)	Dynamin-1 (IPR027741)
ABI family (IPR028457)	Carbonic anhydrase, alpha-class (IPR023561)
Triacylglycerol lipase family (IPR000734)	Cornichon (IPR003377)
Metal-dependent protein hydrolase (IPR003226)	T-complex 11 (IPR008862)
Small heat shock protein HSP20 (IPR031107)	Protein phosphatase 2C family (IPR015655)
S-adenosylmethionine synthetase (IPR002133)	Sirtuin family (IPR003000)
Cancer susceptibility candidate protein 1 (IPR023247)	Ribosomal protein S15 (IPR000589)
Ubiquinone biosynthesis protein COQ9 (IPR012762)	Beta-sarcoglycan (IPR027659)
Ribosomal protein S28e (IPR000289)	Sarcoglycan complex subunit protein (IPR006875)
3'5'-cyclic nucleotide phosphodiesterase	Peptidoglycan recognition protein (IPR015510)
Protein of unknown function DUF3588 (IPR021987)	Lon protease, bacterial/eukaryotic-type (IPR004815)
Phosphotransferase KptA/Tpt1 (IPR002745)	Dynamin superfamily (IPR022812)
Peptidase M10A (IPR021190)	---

Table VI.5. Cellular component-related Gene Ontology (GO) annotations and the corresponding number of transcripts associated with each cellular component GO term identified for upregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts using Blast2GO.

Upregulated Cellular Component GO	Number of Transcripts
Nucleus	2
Cytoplasm	1
RNA-directed RNA polymerase complex	1
Fatty acid synthase complex	1
Extracellular region	1
Ribosome	1
Tubulin complex	1
Extracellular matrix	1

Table VI.6. Cellular component Gene Ontology (GO) annotations and the corresponding number of transcripts associated with each cellular component GO term identified for downregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts using Blast2GO.

Cellular Component GO	Number of Transcripts
RNA-directed RNA polymerase complex	4
Integral component of membrane	4
Viral capsid	3
Membrane	3
Host cell membrane	3
Virion	1
Protein serine/threonine phosphatase complex	1
Transcription factor complex	1
Ribosome	1
Sarcoglycan complex	1
Nucleus	1
Intracellular	1

Table VI.7. Molecular function Gene Ontology (GO) annotations and the corresponding number of transcripts associated with each molecular function GO term identified for upregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts using Blast2GO.

Upregulated Molecular Function GO	Number of Transcripts
Protein binding	4
Metal ion binding	3
Zinc ion binding	3
Nucleic acid binding	2
Hydrolase activity	2
Metalloendopeptidase activity	1
Binding	1
Chitin binding	1
3'5'-cyclic-nucleotide phosphodiesterase activity	1
Carboxylic ester hydrolase activity	1
Phosphtransferase activity, alcohol group as acceptor	1
Transferase activity, transferring phosphorus-containing	1
Microtubule biding	1
Fatty acid synthase activity	1
N-acetylmuramoly-L-alanine amidase activity	1
NAD+ ADP-ribosyltransferase activity	1
Methionine adenosyltransferase activity	1
ATP binding	1
Lysozyme activity	1
RNA-directed RNA polymerase activity	1
Protein dimerization activity	1
RNA binding	1
Stuctural constituent of ribosome	1

Table VI.8. Molecular function Gene Ontology (GO) annotations and the corresponding number of transcripts associated with each molecular function GO term identified for downregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts using Blast2GO.

Downregulated Molecular Function GO	Number of Transcripts
RNA-directed RNA polymerase activity	4
ATP binding	4
RNA binding	4
Nucleic acid binding	4
RNA helicase activity	3
Cysteine-type endopeptidase activity	3
Transferase activity	3
Structural molecule activity	3
Metal ion binding	2
Heterocyclic compound binding	2
Binding	2
Organic cyclic compound binding	2
Xinc ion binding	2
Peptidase activity	2
Nucleotide binding	2
Transcription factor activity	1
Aspartic-type endopeptidase activity	1
N,N-dimethylaniline monooxygenase activity	1
Cystein-type peptidase activity	1
Calcium ion binding	1
Hydrolase activity	1
NAD+ binding	1
Serine-type endopeptidase activity	1
Helicase activity	1
ATP-dependent peptidase activity	1
Hydrolase activity, acting on ester bonds	1
Protein bidning	1
NADP binding	1
GTPase activity	1
Protein serine/threonine phosphatase activity	1
Sequence-specific DNA binding	1
Oxidoreductase activity	1
Flavin adenine dinucleotide binding	1
N-acetylmuramoyl-L-alanine amidase activity	1
Structural constituent of ribosome	1

Table VI.9. Biological process Gene Ontology (GO) annotations and the corresponding number of transcripts associated with each biological process GO term identified for upregulated and downregulated (FDR <0.05) *Spiroplasma kunkelii*-infected *Exitianus exitiosus* transcripts using Blast2GO.

Upregulated Biological Process GO	Number of Transcripts	Downregulated Biological Process GO	Number of Transcripts
Carboydrate metabolic process	2	Proteolysis	6
Purine nucleobase metabolic process	2	Purine nucleobase metabolic process	4
RNA splicing, via endonucleolytic cleavage activity	1	Viral RNA genome replication	4
Metabolic process	1	Transcription, DNA-templated	4
Fatty acid biosynthetic process	1	Oxidation-reduction process	3
Primary metabolic process	1	Transcription, RNA-templated	3
Defense response	1	RNA metabolic process	2
Microtubule cytoskeleton organization	1	Intracellular signal transduction	1
Peptidoglycan catabolic process	1	DNA integration	1
Viral genome replication	1	Transport	1
Chitin metabolic process	1	Peptidoglycan biosynthetic process	1
Proteolysis	1	Clathrin-mediated endocytosis	1
Ubiquinone biosynthetic process	1	Muscle organ development	1
Signal transduction	1	Regulation of transcription, DNA-templated	1
Cytokinesis	1	Obsolete ATP-dependent proteolysis	1
S-adenosylmethionine biosynthetic process	1	Peptidoglycan catabolic process	1
Acyl-carrier-protein biosynthetic process	1	Transmembrane transport	1
Peptidoglycan biosynthetic process	1	Biosynthetic process	1
Regulation of transcription, DNA-templated	1	Protein catabolic process	1
Macromolecule metabolic process	1	Translation	1
Translation	1	Protein dephosphorylation	1
Ribosome biogenesis	1	Ribosome biogenesis	1
Trancription, RNA-templated	1	---	---

upregulated. Several transcripts (7 total) with homology to *Graminella nigrifrons* (black-faced leafhopper) virus 1 (E-values in the range $2.18E^{-45}$ – $1.37E^{-6}$) (Accession YP_009129265) were significantly downregulated.

Immune response and defense related genes were identified in both the upregulated and downregulated transcripts. Upregulated immune-related protein transcripts included hemocytin, an insect humoral lectin (Kotani et al., 1995), a defensin-like protein, which is an induced host defense peptide (Hoffmann and Hetru, 1992), and an i-type (invertebrate) lysozyme (Callewaert and Michiels, 2010) (Table VI.2). Immune-related transcripts were also downregulated, including holotricin 1, which has antibacterial activity against Gram-positive bacteria (Lee et al., 1995) and an MD-2-related lipid-recognition protein, which is part of the Toll pathway (Shimazu et al., 1999) (Table VI.3).

Discussion

The transcriptomic differences between *S. kunkelii*-infected and *S. kunkelii*-naïve *E. exitiosus* have been explored in this study. Plant pathogenic spiroplasma species have been demonstrated as having different effects, both positive and negative, on their insect vectors; however, little is known about the genetic mechanisms behind these effects. To my knowledge, this is the first study that takes an RNA-seq approach to uncovering the effects of a phytopathogenic spiroplasma infection in an insect vector. Differential gene regulation, both upregulation and downregulation, in response to *S. kunkelii* infection in *E. exitiosus* is reported herein.

A total of 308 transcripts were significantly differentially regulated (FDR < 0.05). Of these, 42.9% (132 transcripts), had significant blastx hits (E-value < $1x10^{-3}$) and 15.3% (47

transcripts) had GO annotations. Several of these assembled transcripts had homology to viruses (upregulated and downregulated). When these (9 total) were removed from the blastx- and GO-annotated transcripts in order to focus only on likely leafhopper-derived transcripts, 41.1% and 12.7%, respectively, had annotations. Though these percentages may appear low, they were within range of the two other reported *de novo* assembled and annotated Cicadellid transcriptomes. The *G. nigrifrons* transcriptome reported by Chen et al. (2012) annotated approximately 34% of transcripts using the Swiss-Prot database and 12% with GO terms, while 45% of the *Homalodisca vitripennis* (glassy winged sharpshooter) transcriptome predicted proteins were annotated using tblastx (E-value $\leq 1E^{-5}$) and the NCBI nr (non-redundant) database (Nandety et al., 2013). Though the percentages reported here represent only the significantly differentially regulated transcripts, these numbers will likely prove to reflect those of all assembled transcripts.

One area of extensive study regarding spiroplasma-insect relationships has been assessing the insect host immune response to spiroplasma infection. In examining insect immune responses to invading pathogens, many previous studies have evaluated samples at various intervals relatively soon after exposure; for example, *Drosophila* immune response was evaluated at 24 hours and 5 days after exposure to *S. citri*, and *C. haematoceps* was evaluated at time points ranging from 30 minutes to 6 days after exposure to *S. citri* (Herren and Lemaitre, 2011; Eliautout et al., 2016). In this study, RNA was extracted 26 days after first exposure. Infection at this point was well established in the insect, and differentially regulated defense responses observed were likely sustained and ongoing. RNA was extracted at 26 days post-exposure only, therefore response could not be evaluated over successive time intervals as in aforementioned studies; however, the differential regulation of defense-related genes could indicate a persistent response to spiroplasma virulence. The negative effects of phytopathogenic spiroplasma infection on experimental vector fitness (Madden and Nault, 1983) attest to the pathogenesis of spiroplasma

infection within these vectors. However unsuccessful it may be in overcoming the infection, the vector likely launches and maintains a defense response to the invading spiroplasma. Silencing these genes may result in the leafhopper vector succumbing to spiroplasma pathogenicity more quickly, indicating that an immune response is actively defending the leafhopper against spiroplasma attack. Additionally, the seemingly prolonged immune response observed in this study could be reprentative of a typical non-primary vector/pathogenic spiroplasma type of relationship. It would be informative to compare the response of a primary vector, such as *Dalbulus maidis*, to *S. kunkelii* infection.

Insect defense mechanisms activated by invading microorganisms are encompassed by the innate immune system, which includes responses regulated by the Toll, Imd, Jak-STAT, autophagy, and RNA interference pathways (Hoffmann and Reichhart, 2002; Kingslover et al., 2013). Specifically, the Toll pathway controls the response to Gram-positive bacteria as well as fungi. The discovery of a differentially regulated Toll pathway related protein (MD-2-related lipid-recognition protein) transcript could be evidence of similar type of defense response that is activated by invading spiroplasmas, which are thought to have evolved from Gram-positive bacteria (Fox et al., 1980). This is interesting, however, because the Toll cascade is activated by recognition of cell-wall components, which are lacking in (cell wall-less) spiroplasmas; however, MD-2-related is a lipid-recognition protein, which was, curiously, significantly downregulated. Another downregulated immune-related gene included holotricin 1, which also has activity against Gram-positive bacteria, though the specific action is yet unknown. The upregulation of an i-type (invertebrate) lysozyme-related transcript was another interesting discovery. Lysozyme is a hydrolytic enzyme that cleaves peptidoglycan bonds, which are components of the bacterial cell wall. The differential regulation of these bacterial invasion defense response-related genes perhaps indicates an immune response to spiroplasma infection that is homologous to insect bacterial defense response. RNA-seq based analysis of leafhopper transcriptional response to

phytopathogenic spiroplasma exposure and infection over time could further elucidate the immediate and sustained defense responses that occur in vectors.

One of the primary findings by Eliautout et al. (2016) was that hexamerin was required for immune response and survival of *C. haematoceps* after acquisition of *S. citri*. Hexamerin was not observed in either differentially regulated transcript group (upregulated or downregulated; FDR < 0.05); however, after expanding the FDR threshold to 0.10, one hexamerin transcript (E-value: $1x10^{-26}$) was discovered with approximately 10-fold upregulation (FDR < 0.07), indicating that hexamerin protein is likely involved in *E. exitiosus* response to *S. kunkelii* infection as well. Differential regulation was reported herein at a false discovery rate threshold of 0.05 (5%) to maintain high statistical support of differential regulation discoveries. Increasing the FDR threshold to 0.10 (10%) doubles the chance of false discovery compared to the initial threshold; however, likely differential transcript regulation at this level increases from 308 to 468 transcripts. Future analysis of these additional 160 transcripts will likely uncover more biologically relevant differentially expressed defense response genes.

This study provides a first look at a full transcriptomic response in a leafhopper vector to phytopathogenic spriroplasma infection. Defense related genes were activated and differentially expressed, which, until recently (Eliautout et al., 2016), was not thought to occur in spiroplasma/insect host systems. Viral transcripts were also differentially expressed. In particular, a virus homologous to *Graminella nigrifrons* virus 1, a novel Iflavirus recently discovered via virus defense response trancriptome sequencing, (Chen et al., 2015) was downregulated, perhaps indicating that the activated immune response was suppressing transcription of particular viruses. Annotations of differentially expressed proteins and candidate ORFs were lacking; however, with limited annotated insect genome and transcriptome sequences available, this is a common problem for analysis of genome sequences of insects and other eukaryotes (Moran, 2010;

International Aphid Genomics Consortium, 2010; Moreton et al., 2015). Further analysis, such as the annotation of the additional 160 differentially regulated transcripts and experimentation and response anaylsis using both primary and experimental vectors of spiroplasmas, could provide more insights into the *Spiroplasma*-leafhopper vector relationship interface. Novel transmission prevention or disease management strategies will surely be developed from the elucidation of these complex and intricate relationships.

REFERENCES

Alivizatos, A. S. 1982. Feeding behavior of the spiroplasma vectors *Dlbulus maidis* and *Euscelidius variegatus in vivo* and *in vitro*. Chronika Benaki Phytopathological Institutu 13:128-44.

Anbutsu, H. and Fukatsu, T. 2010. Evasion, suppression and tolerance of Drosophila innate immunity by a male-killing Spiroplasma endosymbiont. Insect Molecular Biology 19:481–488

Bandi, C., Dunn, A.M., Hurst, G.D. and Rigaud, T. 2001. Inherited microorganisms, sex-specific virulence and reproductive parasitism. Trends in parasitology 17(2):88-94.

Barros, T. S., Davis, R. E., Resende, R. O., and Dally, E. L. 2001. Design of a polymerase chain reaction for specific detection of corn stunt spiroplasma. Plant Disease, 85(5):475-80.

Callewaert, L. and Michiels, C.W., 2010. Lysozymes in the animal kingdom. Journal of biosciences 35(1):127-160.

Carloni, E., Virla, E., Paradell, S., Carpane, P., Nome, C., Laguna, I. and Pecci, M.G. 2011. Exitianus obscurinervis (Hemiptera: Cicadellidae), a new experimental vector of Spiroplasma kunkelii. Journal of economic entomology 104(6):1793-1799.

Chen, Y., Cassone, B.J., Bai, X., Redinbaugh, M.G. and Michel, A.P. 2012. Transcriptome of the plant virus vector Graminella nigrifrons, and the molecular interactions of Maize fine streak rhabdovirus transmission. PLoS One 7(7):e40613.

Chen, Y., Liu, S., and Bonning, B.C. 2015. Genome sequence of a novel iflavirus from the leafhopper Graminella nigrifrons. Genome announcements 3(2):e00323-15.

Conesa, A., Götz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-76.

Ebbert, M.A. and Nault, L.R. 1994. Improved overwintering ability in Dalbulus maidis (Homoptera: Cicadellidae) vectors infected with Spiroplasma kunkelii (Mycoplasmatales: Spiroplasmataceae). Environmental Entomology 23(3):634-644.

Eliautout, R., Dubrana, M.P., Vincent-Monégat, C., Vallier, A., Braquart-Varnier, C., Poirié, M., Saillard, C., Heddi, A., and Arricau-Bouvery, N. 2016. Immune response and survival of Circulifer haematoceps to Spiroplasma citri infection requires expression of the gene hexamerin. Developmental & Comparative Immunology 54(1):7-19.

Fletcher, J., Wayadande, A., Melcher, U. and Ye, F. 1998. The phytopathogenic mollicute-insect vector interface: a closer look. Phytopathology 88(12):1351-1358.

Fox, G.C.A., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., and Zablen, L.B. 1980. The phylogeny of prokaryotes. Science (New York, NY) 209(4455):457.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29(7):644-652.

Herren, J.K. and Lemaitre, B. 2011. Spiroplasma and host immunity: activation of humoral immune responses increases endosymbiont load and susceptibility to certain Gram-negative bacterial pathogens in Drosophila melanogaster. Cellular microbiology 13(9):1385-1396.

Hoffmann, J.A. and Hetru, C. 1992. Insect defensins: inducible antibacterial peptides. Immunology today 13(10):411-415.

Hoffmann, J.A. and Reichhart, J.M., 2002. Drosophila innate immunity: an evolutionary perspective. Nature immunology 3(2):121-126.

Hogenhout, S.A., Ammar, E., Whitfield A.E., Redinbaugh, M.G. 2008. Insect Vector Interactions with Persistently Transmitted Viruses. Annual Review of Phytopathology 2008.46:327-359.

Hurst, G.D., Anbutsu, H., Kutsukake, M. and Fukatsu, T. 2003. Hidden from the host: Spiroplasma bacteria infecting Drosophila do not cause an immune response, but are suppressed by ectopic immune activation. Insect molecular biology 12(1):93-97.

International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol, 8(2):e1000313.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., and Pesseat, S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236-1240.

Kingsolver, M.B., Huang, Z. and Hardy, R.W. 2013. Insect antiviral innate immunity: pathways, effectors, and connections. Journal of molecular biology 425(24):4921-4936.

Kotani, E., Yamakawa, M., Iwamoto, S.I., Tashiro, M., Mori, H., Sumida, M., Matsubara, F., Taniai, K., Kadono-Okuda, K., Kato, Y. and Mori, H. 1995. Cloning and expression of the gene of hemocytin, an insect humoral lectin which is homologous with the mammalian von Willebrand factor. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1260(3):245-258.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10(3):R25.

Lee, S., Moon, H., Kawabata, S., Kurata, S., Natori, S., and Lee, B. 1995. A sapecin homologue of Holotrichia diomphalia: purification, sequencing and determination of disulfide pairs. Biological and Pharmaceutical Bulletin 18.3:457-459.

Li, B. and Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12(1):1

Maramorosch, K. and Harris, K. ed., 1979. Leafhopper vectors and plant disease agents. Elsevier.

Marçais, G. and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764-770

Madden, L.V. and Nault, L.R. 1983. Differential pathogenicity of corn stunting mollicutes to leafhopper vectors in Dalbulus and Baldulus species. Phytopathology 73(12):1608-1614.

Madden, L.V., Nault, L.R., Heady, S.E. and Styer, W.E. 1984. Effect of maize stunting mollicutes on survival and fecundity of Dalbulus leafhopper vectors. Annals of applied biology 105(3):431-441.

Maramorosch, K. 1981. Spiroplasmas: agents of animal and plant diseases. BioScience 31(5):374-380.

Moran, M.A., 2010. Metatranscriptomics: eavesdropping on complex microbial communities. *Issues*.

Moreton, J., Izquierdo, A. and Emes, R.D. 2015. Assembly, Assessment, and Availability of De novo Generated Eukaryotic Transcriptomes. Frontiers in Genetics, 6.

Nandety, R.S., Kamita, S.G., Hammock, B.D. and Falk, B.W. 2013. Sequencing and de novo assembly of the transcriptome of the glassy-winged sharpshooter (Homalodisca vitripennis). PloS one 8(12):e81681.

Nault, L. R. 1980. Maize bushy stunt and corn stunt: a comparison of disease symptoms, pathogen host ranges, and vectors. Phytopathology 70:659-662.

O'Hayer, K.W., Schultz, G.A., Eastman, C.E., Fletcher, J., and Goodman, R.M., 1983. Transmission of Spiroplasma citri by the aster leafhopper, Macrosteles fascifrons (Homoptera: cicadellidae). Annals of Applied Biology 102:311-318.

Ossiannilsson, F. 1966. Insects in the epidemiology of plant viruses. Annual review of entomology 11(1):213-232.

Purcell AH. Insect vector relationships with procaryotic plant pathogens. Annual Review of Phytopathology. 1982;20:397–417.

Regassa, L.B. and Gasparich, G.E. 2006. Spiroplasmas: evolutionary relationships and biodiversity. Frontiers in Bioscience 11:2983-3002.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:1.

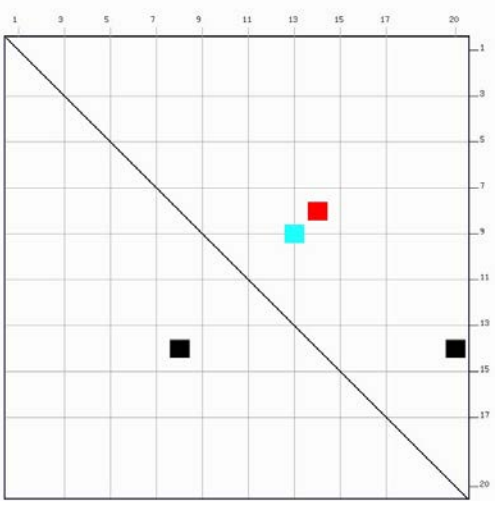
Shimazu, R., Akashi, S., Ogata, H., Nagai, Y., Fukudome, K., Miyake, K. and Kimoto, M. 1999. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. The Journal of experimental medicine 189(11):1777-1782.

Weintraub, P. G. and Beanland, L. 2006. Insect vectors of phytoplasmas. Annual Review of Entomology 51:91-111.

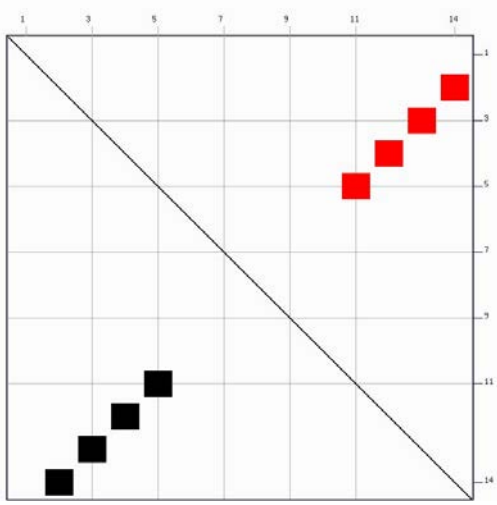
APPENDICES

Appendix 1. mFold evaluations of each novel *Bemisia tabaci* biotype- and *Trialeurodes vaporariorum* species-specific primer.

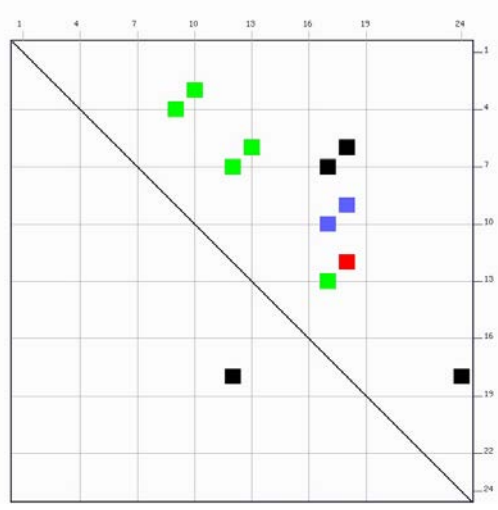
QF: ΔG = 0.9 kcal/mol



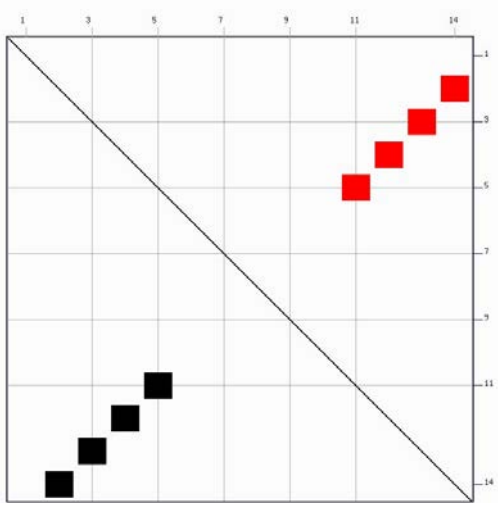
QR: ΔG = 0.0 kcal/mol



QF-f: ΔG = 1.0 kcal/mol

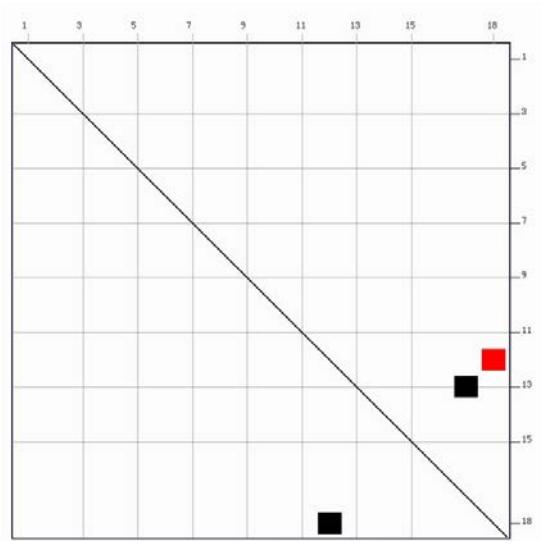


QR: ΔG = 0.0 kcal/mol

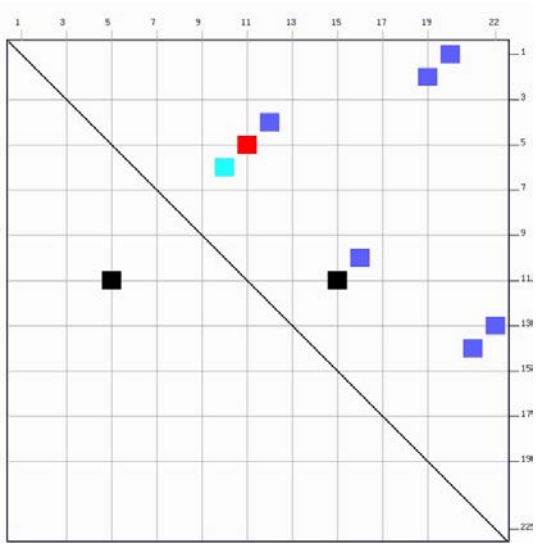


Appendix 1 (cont.). mFold evaluations of each novel *Bemisia tabaci* biotype- and *Trialeurodes vaporariorum* species-specific primer.

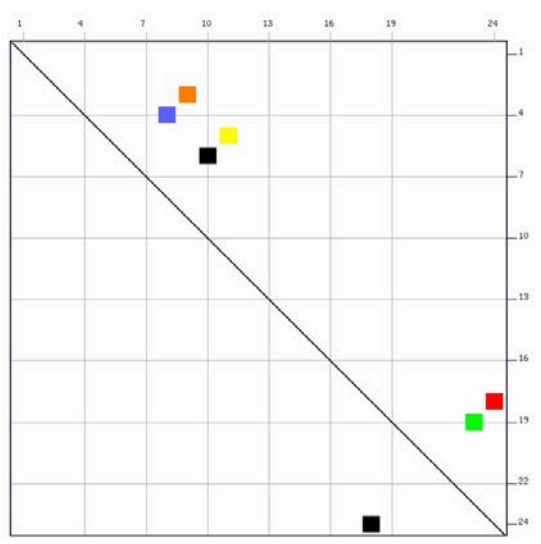
BF: ΔG = 0.3 kcal/mol



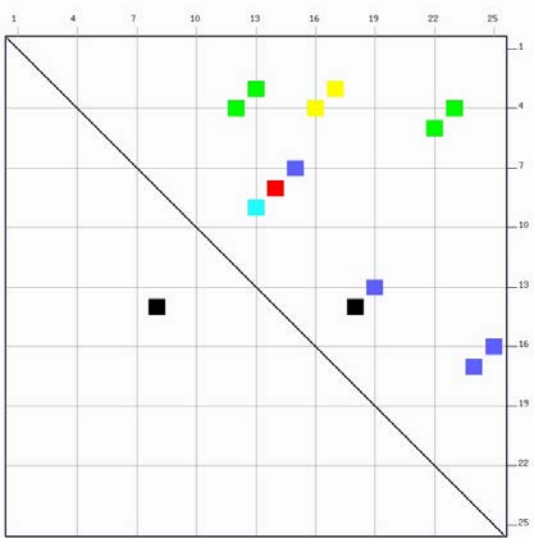
BR: ΔG = 1.0 kcal/mol



BF-f: ΔG = 0.7 kcal/mol

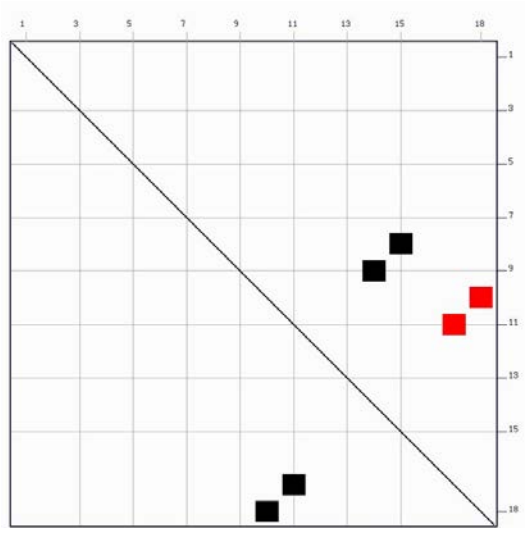


BR-f: ΔG = 1.0 kcal/mol

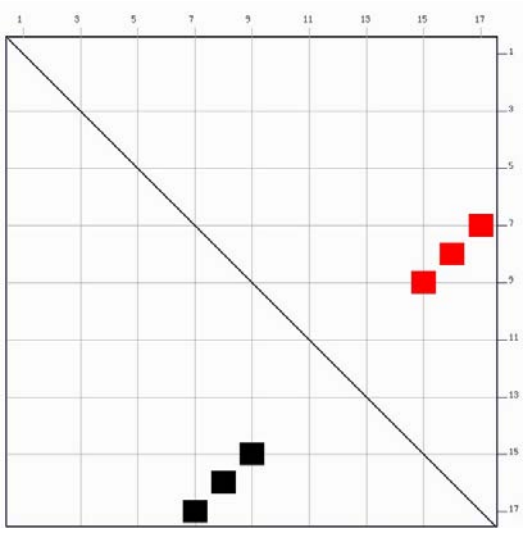


Appendix 1 (cont.). mFold evaluations of each novel *Bemisia tabaci* biotype- and *Trialeurodes vaporariorum* species-specific primer.

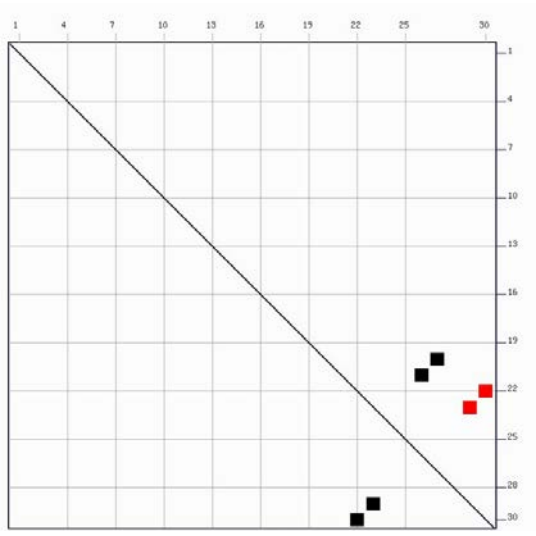
AF: ΔG = 0.9 kcal/mol



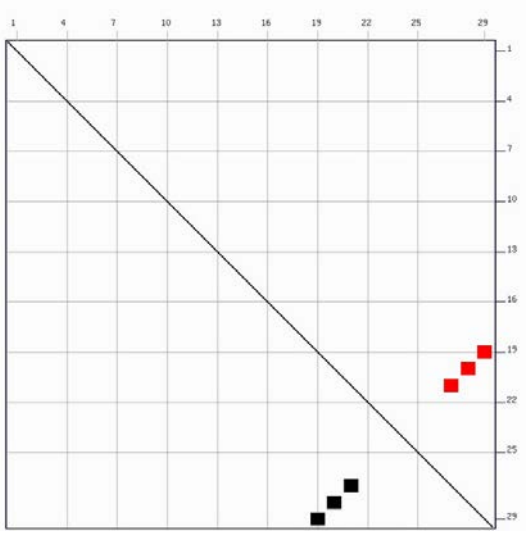
AR: ΔG = 0.0 kcal/mol



AF: ΔG = 0.9 kcal/mol

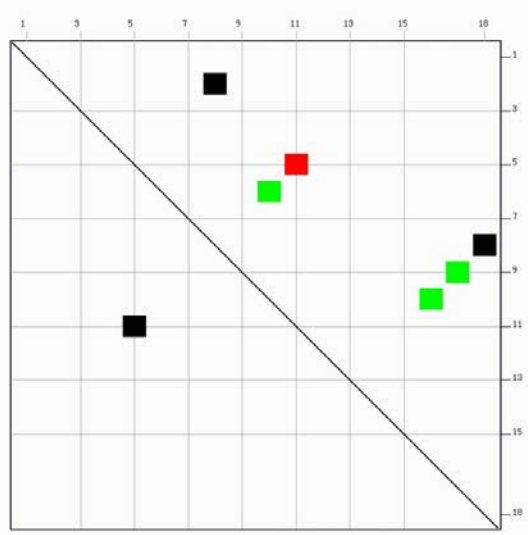


AR: ΔG = 0.0 kcal/mol

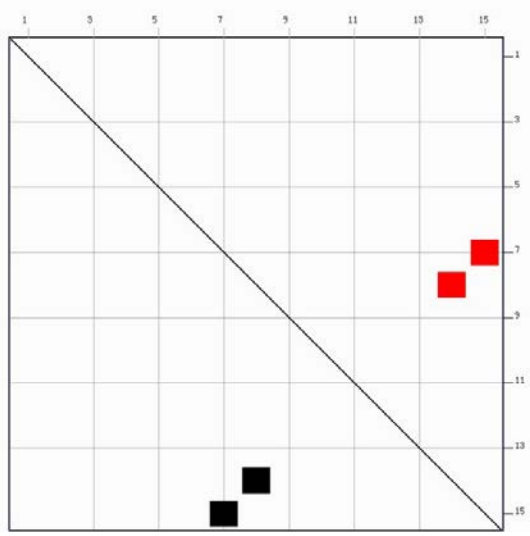


Appendix 1 (cont.). mFold evaluations of each novel *Bemisia tabaci* biotype- and *Trialeurodes vaporariorum* species-specific primer.

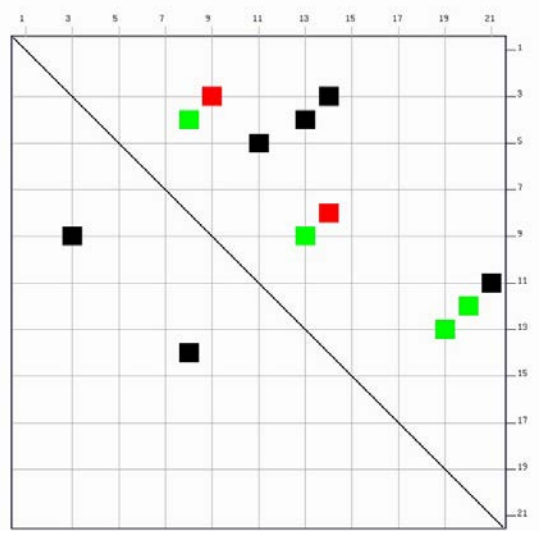
TvF: ΔG = 1.0 kcal/mol



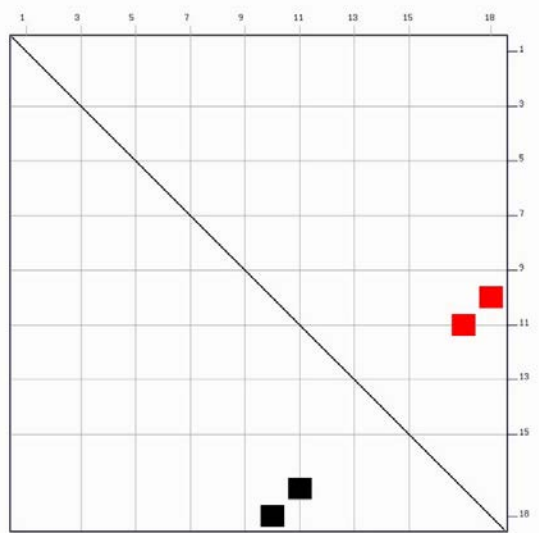
TvR: ΔG = 0.0 kcal/mol



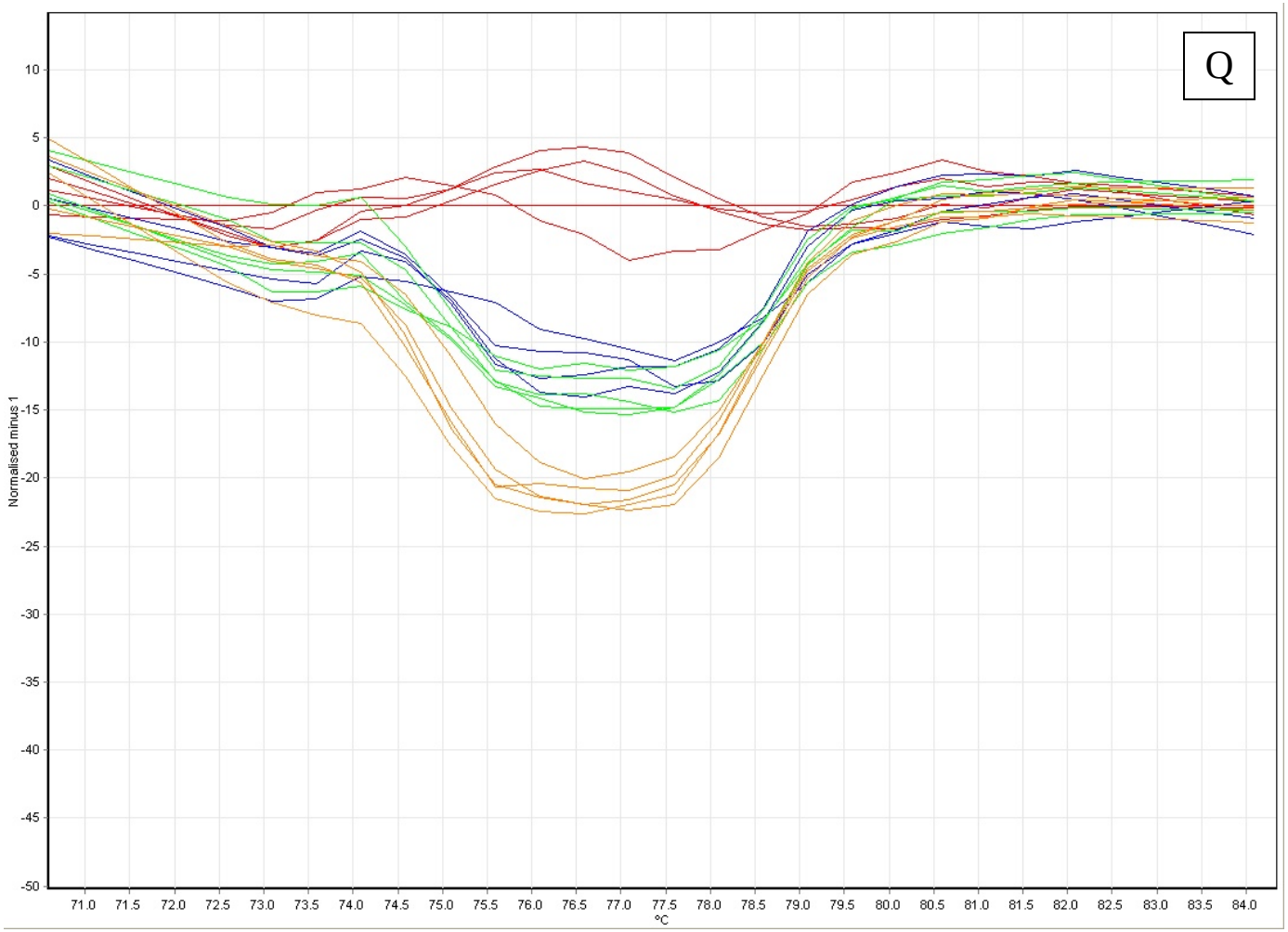
TvF-f: ΔG = 1.0 kcal/mol



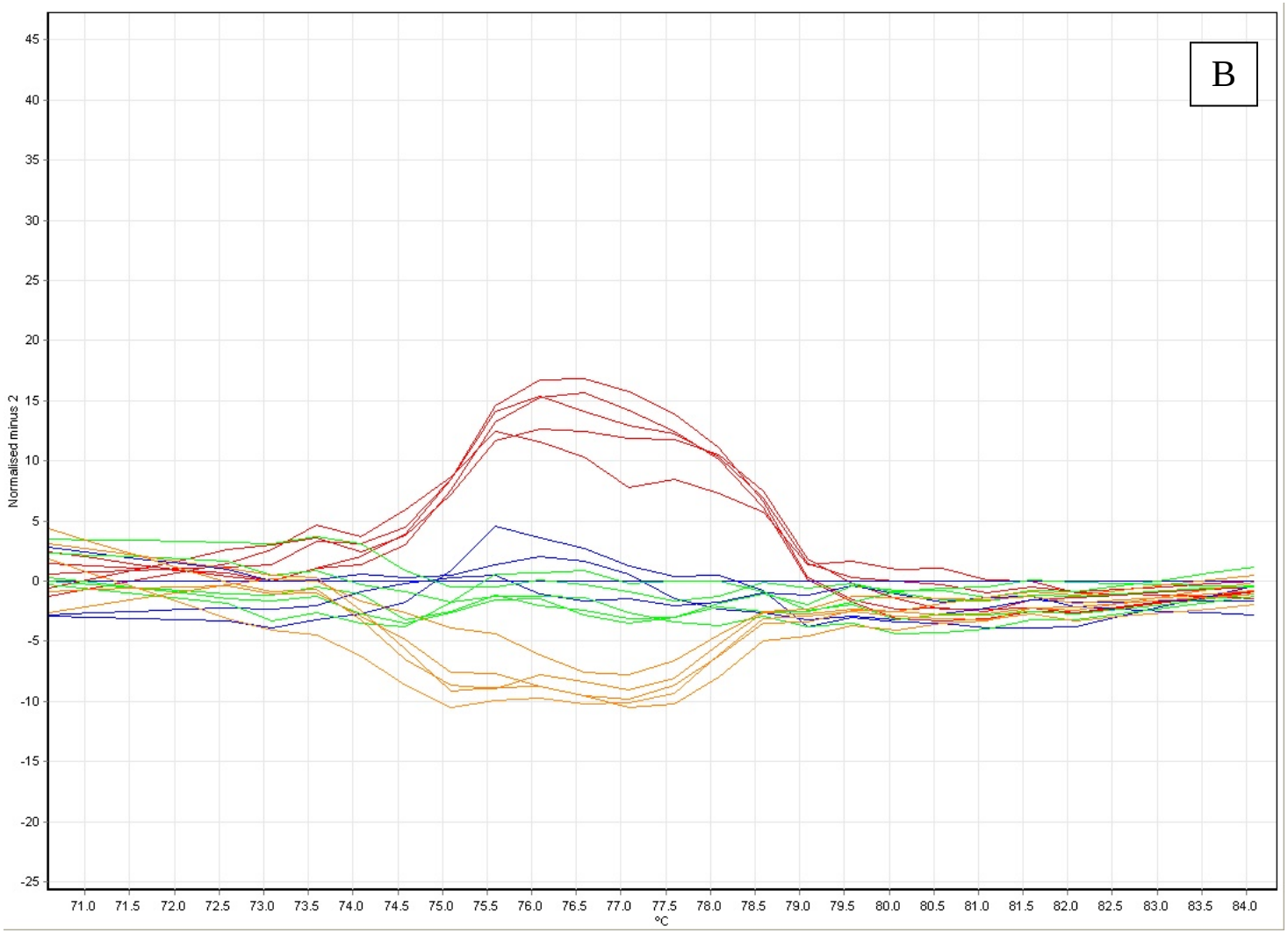
TvR-f: ΔG = 0.0 kcal/mol



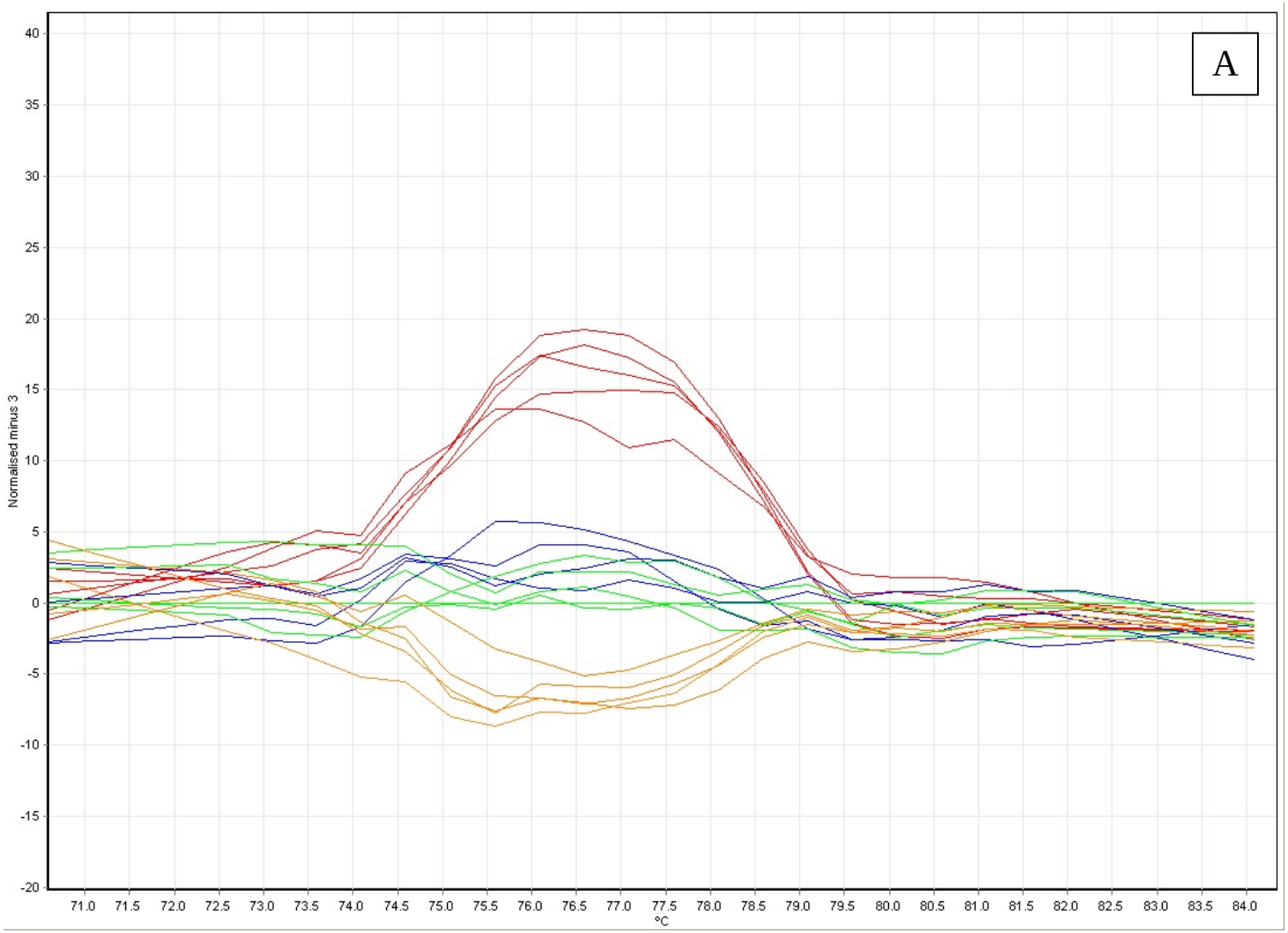
Appendix 2. Difference graphs plotted against control genotypes assayed by MTA with *B. tabaci* biotype and *T. vaporariorum* primer sets. Graphs are in the order Q biotype (red), B (blue), A (green), and *T. vaporariorum* (orange).



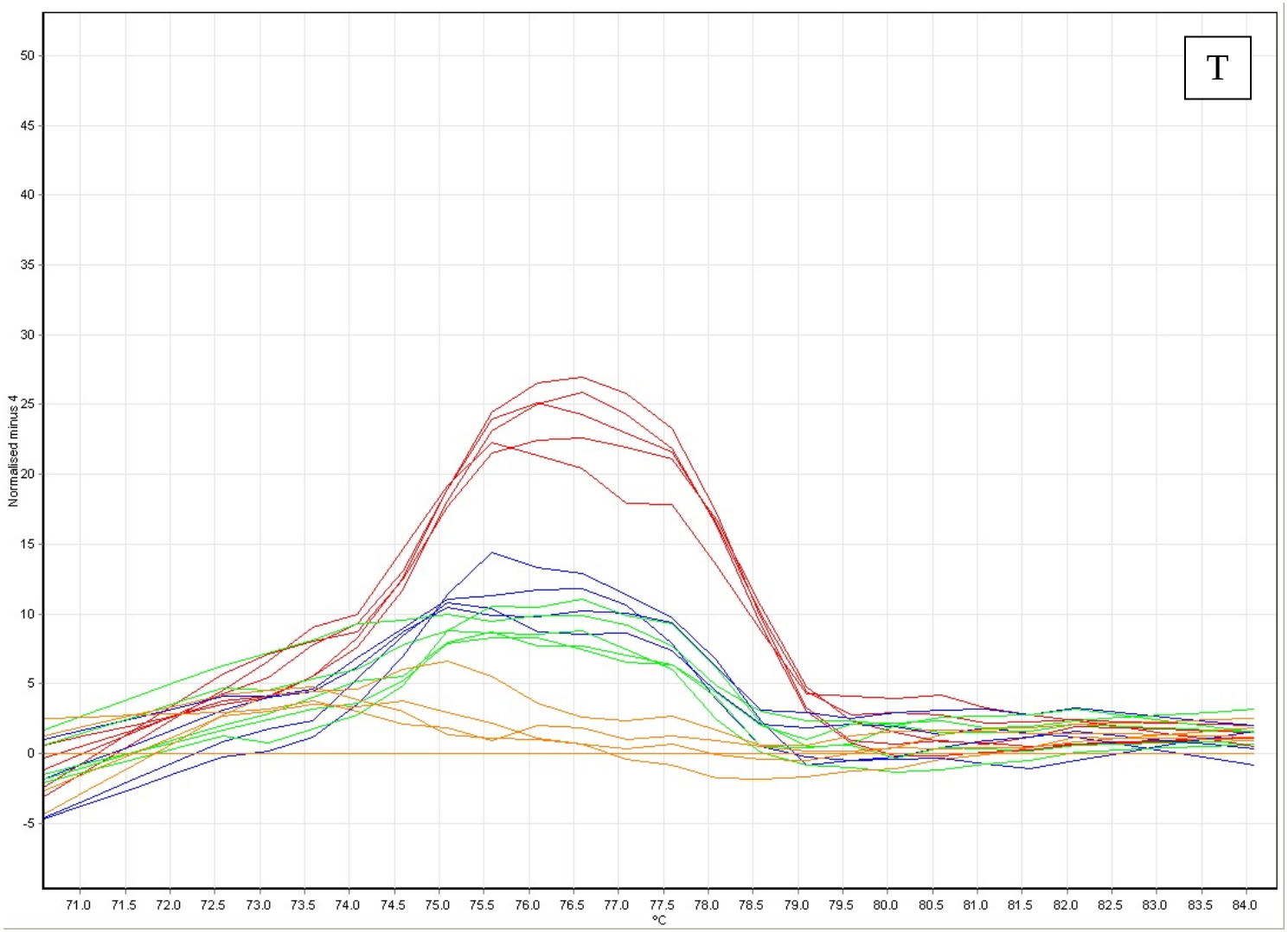
Appendix 2 (cont.). Difference graphs plotted against control genotypes assayed by MTA with *B. tabaci* biotype and *T. vaporariorum* primer sets. Graphs are in the order Q biotype (red), B (blue), A (green), and *T. vaporariorum* (orange).



Appendix 2 (cont.). Difference graphs plotted against control genotypes assayed by MTA with *B. tabaci* biotype and *T. vaporariorum* primer sets. Graphs are in the order Q biotype (red), B (blue), A (green), and *T. vaporariorum* (orange).



Appendix 2 (cont.). Difference graphs plotted against control genotypes assayed by MTA with *B. tabaci* biotype and *T. vaporariorum* primer sets. Graphs are in the order Q biotype (red), B (blue), A (green), and *T. vaporariorum* (orange).



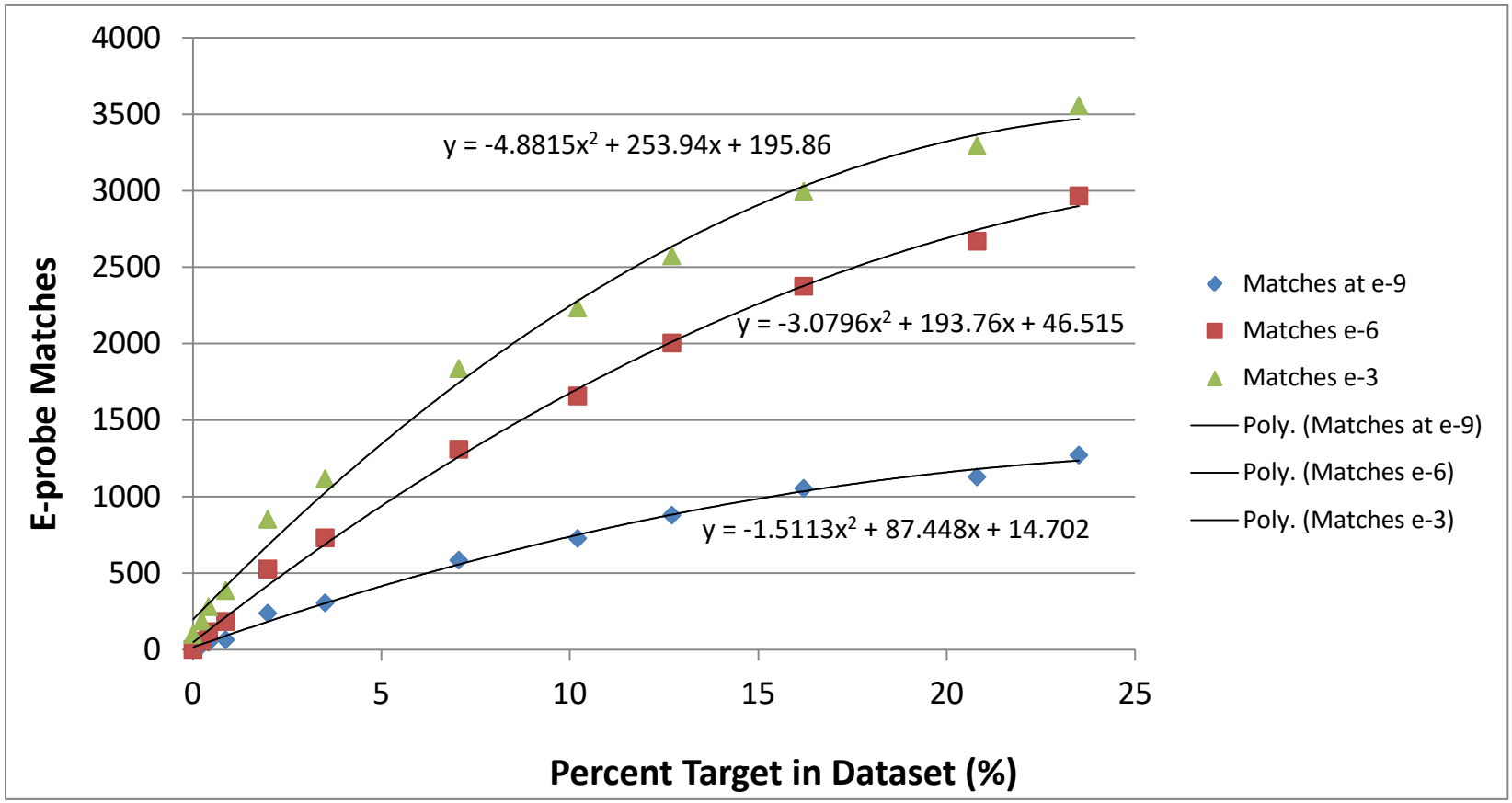
Appendix 3. Results of high (H), medium (M), low (L), and very low (VL) target (*Spiroplasma kunkelii*) percentage mock sample databases (MSDs) query with *S. kunkelii*-specific e-probes and decoy e-probes parsed at E-values of 10^{-9}, 10^{-6}, and 10^{-3}.

MSD	Percent Target	E-value	E-probe Matches	E-probe Hits	Decoy Matches	Decoy Hits	Chi-square	P-value
H1	16.21	e^{-9}	1053	1988	0	0	1170.03	< 0.00001
H2	20.81		1128	2339	0	0	1263.36	< 0.00001
H3	23.51		1271	2720	0	0	1445.51	< 0.00001
M1	12.71		877	1517	0	0	956.69	< 0.00001
M2	10.21		725	1057	0	0	778.62	< 0.00001
M3	7.06		583	884	0	0	617.18	< 0.00001
L1	3.51		305	382	0	0	314.10	< 0.00001
L2	1.99		237	269	0	0	242.46	< 0.00001
L3	0.87		63	64	0	0	63.38	< 0.00001
VL1	0.24		29	29	0	0	29.08	< 0.00001
VL2	0.01		0	0	0	0	--	--
VL3	0.42		46	46	0	0	46.20	< 0.00001
H1	16.21	e^{-6}	2375	4381	4	4	3052.88	< 0.00001
H2	20.81		2670	5474	0	0	3577.22	< 0.00001
H3	23.51		2965	6311	0	0	4127.40	< 0.00001
M1	12.71		2004	3516	4	8	2461.52	< 0.00001
M2	10.21		1657	2514	0	0	1966.51	< 0.00001
M3	7.06		1308	1948	0	0	1493.56	< 0.00001
L1	3.51		730	930	4	4	771.90	< 0.00001
L2	1.99		526	585	4	4	541.38	< 0.00001
L3	0.87		183	184	0	0	186.24	< 0.00001
VL1	0.24		55	55	0	0	55.29	< 0.00001
VL2	0.01		1	1	0	0	1.000	0.31731
VL3	0.42		115	115	0	0	116.27	< 0.00001
H1	16.21	e^{-3}	2997	7073	125	170	3755.76	< 0.00001
H2	20.81		3292	9163	121	160	4359.40	< 0.00001
H3	23.51		3557	10059	153	195	4822.74	< 0.00001
M1	12.71		2573	5838	102	138	3060.07	< 0.00001
M2	10.21		2233	4408	120	149	2443.63	< 0.00001
M3	7.06		1836	3329	115	133	1863.44	< 0.00001
L1	3.51		1117	1656	90	101	987.00	< 0.00001
L2	1.99		850	1071	110	126	627.65	< 0.00001
L3	0.87		385	431	89	105	193.56	< 0.00001
VL1	0.24		184	208	90	104	33.11	< 0.00001
VL2	0.01		99	136	102	126	0.046	0.83017
VL3	0.42		279	303	102	125	85.32	< 0.00001

Appendix 4. Averages of high (H), medium (M), low (L), and very low (VL) target (*Spiroplasma kunkelii*) percentage mock sample databases (MSDs) queried with *S. kunkelii*-specific e-probes and decoy e-probes parsed at E-values of 10^{-9}, 10^{-6}, and 10^{-3}.

MSD	Average Target (%)	Parser E-value	Average Matches	Average Hits	Decoy Matches	Decoy Hits	Chi-square	P-value
H	20.18	e^{-9}	1150.67	2349.00	0	0	1291.87	< 0.00001
M	9.99		728.33	1152.67	0	0	782.46	< 0.00001
L	2.12		201.67	238.33	0	0	205.61	< 0.00001
VL	0.22		25.00	25.00	0	0	25.06	< 0.00001
H	20.18	e^{-6}	2670.00	5388.67	1.33	1.33	3572.48	< 0.00001
M	9.99		1656.33	2659.33	1.33	2.67	1961.13	< 0.00001
L	2.12		479.67	566.33	2.67	2.67	494.369	< 0.00001
VL	0.22		57.00	57.00	0	0	57.31	< 0.00001
H	20.18	e^{-3}	3282.00	8765.00	133	175	4297.82	< 0.00001
M	9.99		2214.00	4525.00	112.33	140	2437.26	< 0.00001
L	2.12		784.00	1052.67	96.33	110.67	586.19	< 0.00001
VL	0.22		187.33	215.67	98	118.33	28.75	< 0.00001

Appendix 5. Scatter plot of MSD query matches parsed at 10^{-9}, 10^{-6}, and 10^{-3} with corresponding polynomial (order: 2) trendlines.



VITA

Sharon Ann Andreason

Candidate for the Degree of

Doctor of Philosophy

Thesis: DISCRIMINATION AND TRANSCRIPTIONAL RESPONSE ANALYSIS OF HEMIPTERAN PHYTOPATHOGEN VECTORS

Major Field: Plant Pathology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in July, 2016.

Completed the requirements for the Bachelor of Science in Biology at University of Texas at Tyler, Tyler, Texas in 2009.

Experience:

Insect vectors, DNA and RNA extraction, PCR, primer design, gel electrophoresis, Sanger sequencing, Illumina sequencing, RNA-Seq, bioinformatics, insect colony rearing, insect salivary gland dissection, spiroplasma culturing, dark-field microscopy, light microscopy, basic molecular and microbiological techniques

Professional Memberships:

American Phytopathology Society
Entomological Society of America